1998

Capillary gel electrophoresis : a tool for DNA separation

Anupma Dubey
San Jose State University

INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps. Each original is also photographed in one exposure and is included in reduced form at the back of the book.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

UMI

A Bell & Howell Information Company
300 North Zeeb Road, Ann Arbor MI 48106-1346 USA
313/761-4700 800/521-0600

CAPILLARY GEL ELECTROPHORESIS:

A TOOL FOR DNA SEPARATION

A Thesis

Presented to

The Faculty of the Department of Chemistry

San Jose State University

In Partial Fulfillment

of the Requirements for the Degree

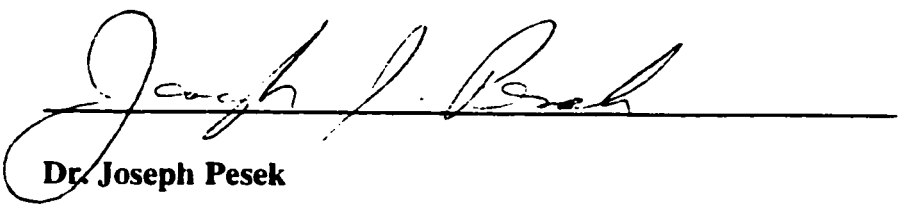
Master of Arts

by

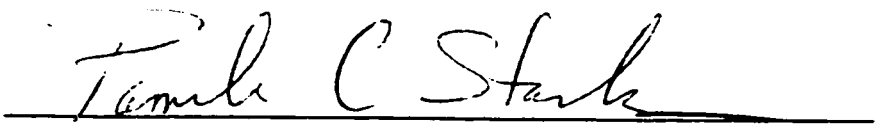
Anupma Dubey

August 1998

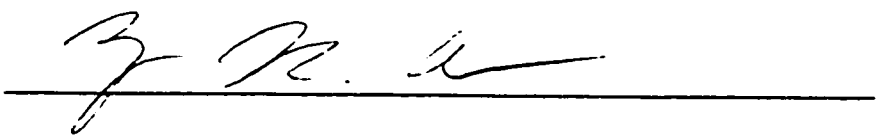
APPROVED FOR THE DEPARTMENT OF CHEMISTRY



Dr. Joseph Pesek



Dr. Pam Stacks



Dr. Roy Okuda

APPROVED FOR THE UNIVERSITY

ABSTRACT

CAPILLARY GEL ELECTROPHORESIS:

A TOOL FOR DNA SEPARATION

by Anupma Dubey

Over the last ten years separations using Capillary Gel Electrophoresis (CGE) and subsequent visualization of nucleic acids, gene structure and nucleotides have improved tremendously. This progress has led to the biggest concerted endeavor in the history of science, i.e. the mapping of the human genome. CGE offers high resolution on the analytical scale.

This paper describes the CGE technique, its application in DNA separation and sequencing and its comparison to conventional slab-gel electrophoresis. The underlying capillary electrophoresis (CE) theory, instrumentation and various factors that affect its performance are described. Several successful applications that have been developed with this technique such as single stranded and double stranded DNA separation, DNA sequencing, polymerase chain reaction (PCR) product analysis, and anti-sense DNA analysis are presented. The paper also discusses the Human Genome Project and the role of CGE in that context.

Acknowledgments

I am greatly thankful to Dr. Joseph Pesek, my research advisor, for his guidance, support and encouragement. Halfway through my thesis preparation, I relocated to Oregon. This presented a lot of logistic problems and hurdles in completing my thesis. At times, I felt like giving up but Dr. Pesek encouraged me to still go on and complete the work I had started. He was very patient and understanding in helping me complete the thesis remotely. I am also thankful to Dr. P. Stacks and Dr. R. Okuda for taking out time to go through my drafts and giving me helpful hints and direction.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

1. Introduction:

The tremendous demands of the Human Genome Project (HGP) for improved DNA sequencing methodologies has increased the search for innovative new approaches to high speed DNA sequence analysis. One of the most promising of these new techniques is the capillary gel electrophoresis (CGE).

CGE has now become a widely accepted analytical tool for the separation and analysis of biopolymers like DNA and proteins. CGE has been used very successfully for the separation of DNA fragments resulting in some of the most efficient separations reported to date.

Potential advantages over conventional electrophoretic methods include high speed separation, high resolution, on-line detection, capability for automation, and detection of small quantities of analyte(s). Figure 1 shows the characteristics[1] of a CE instrument and Figure 2 shows the general schematic of a CE instrument. It consists of a narrow-bore, fused silica capillary, the two ends of which are immersed in reservoirs containing buffer solution. The reservoirs also contain the electrodes used to make electrical connection between the high voltage power supply and the capillary. The sample is loaded on to the capillary by replacing one of the reservoirs (usually at the anode) with the sample reservoir and applying either an electric field or an external pressure. After replacing the buffer reservoir, the electric field is applied and the separation is performed. The gel not only provides the separation matrix, but also acts as an anti-convective media. The high surface-to-volume ratio of the capillary provides an easy way to dissipate the heat

generated by the passage of current. The software for CE systems has also improved significantly over the last couple of years. The software available for the CE systems is capable of programming CE methodology, calculating analyte mobility, connecting peak areas for electrophoretic velocity, and estimating molecular properties like molecular weight.

2. Theory:

2.1 Electrophoresis:

Separation by electrophoresis is based on differences in solute velocity in an electrical field[2]. The velocity of an ion is given by:

$$v = \mu_e E$$

where:

v = ion velocity

μ_e = electrophoretic mobility

E = applied electric field

The electric field is the function of the applied voltage and capillary length (in volts/cm). The mobility, for a certain ion and medium, is a characteristic of that ion and is a constant. Electrophoretic mobility in terms of physical parameters is given by:

$$\mu_e = q/6\pi\eta r$$

where:

q = ion charge

η = solution viscosity

r = ion radius

μ_e = electrophoretic mobility

The above equation shows that small, highly charged species will have higher mobility compared to a large, minimally charged species. The absolute mobility of an ion is the mobility determined at the point of full solute charge and extrapolated to infinite dilution. Effective mobility of an ion is dependent on pH (i.e. solute pKa) and the composition of the running buffer. Figure 3 shows the mobility of two weak acids as a function of pH. Thus two species that may have the same absolute mobility can still be separated because of their different pKa values depending on their pH controlled charge.

2.2 Electroosmotic flow (EOF):

EOF is the bulk flow of a liquid in the capillary because of the surface charge on the interior capillary wall (Figure 4). EOF is a fundamental constituent of CE[3]. When an electric field is applied to the solution, a double layer forms at the wall of the capillary, and EOF results. This flow controls the amount of time the analytes remain in the capillary by superposition of flow onto solute mobility.

Most solid surfaces under aqueous conditions possess an excess of negative charge. This negative charge could be because of ionization of the surface as in a fused silica capillary or from adsorption of ionic species at the surface. Figure 5 shows the development of the EOF. The magnitude of EOF can be expressed in terms of mobility as:

$$\mu_{EOF} = \epsilon \zeta / \eta \quad \text{or} \quad V_{EOF} = (\epsilon \zeta / \eta) E$$

where:

μ_{EOF} = electroosmotic mobility

V_{EOF} = electroosmotic velocity

ζ = zeta potential

ϵ = dielectric constant

Note: Mobility is independent of the applied electric field.

The zeta potential is the potential difference very close to the wall (Figure 5b).

A unique feature of EOF is its flat profile of flow in the capillary (Figure 6). This kind of flow is uniform throughout and therefore does not directly contribute to the dispersion of solute zones. The flow profile generated by an external pump on the other hand, has a laminar or parabolic flow due to the shear force at the wall. Figure 6 also shows the difference between a flat profile flow and laminar flow, and also their corresponding solute zones.

Another benefit of EOF is that it moves nearly all species regardless of charge in the same direction. Therefore cations, anions, and neutral compounds can all be driven through the capillary in a single run (Figure 7).

3. Instrumentation:

3.1 Capillaries:

Ideally a capillary should be chemically and electrically inert, UV-Visible transparent, flexible, robust and inexpensive. Fused silica meets all of the above requirements and hence is the most commonly used material for capillaries. Capillaries are used as an electromigration channel for the separation of many different kinds of analytes, including

biological molecules. The physical characteristics of narrow bore capillaries make them ideal for electrophoresis. Fused silica capillaries used in CGE typically have an internal diameter (i.d.) of 20-100 microns, an outer diameter (o.d.) of 100-400 microns and a length of 20-100 cm (Figure 8). The polyimide coating on the outside makes these fragile capillaries flexible and easy to use. The high surface-to-volume ratio of capillaries allows for very efficient dissipation of joule heat generated from large electric fields. Table I compares the surface-to-volume ratios of a standard analytical slab gel system with typical capillaries used in CE[3]. The small diameter of capillaries also means that very small quantities of sample (nanoliter range) and buffer (milliliter range) are used for the separation process.

3.1.1 Preparation for CE:

A capillary of desired length can be cut using a ceramic knife or an ampoule file, taking care that both ends are square and flat for proper introduction of sample. A window for on-line detection is made by removing the polyimide coating. The coating can be removed by burning with a flame or by scraping off the layer or by using hot (103 °C) H_2SO_4 or hot concentrated KOH. The last method is preferred in case of an internal coating because other methods of polyimide removal might damage part of the coating. The detection window is closer to the outlet and is ideally 0.3 cm in length (not more than 1 cm). Before use, an uncoated fused silica capillary is "preconditioned" by rinsing with 100 mM NaOH followed by H_2O and finally by the separation buffer.

3.1.2 Conditioning:

A new capillary is always equilibrated with the separation buffer. This process is called conditioning and is also used when an old capillary is used for analysis that requires a different separation buffer. Since this equilibration takes several hours, it is a good idea to dedicate certain capillaries to individual buffer systems. This saves time and the storage of capillaries also becomes easy (the capillary can be stored in the buffer it uses for separation for some time between use).

3.1.3 Cleansing and storage:

The capillary needs to be reconditioned or cleansed in the case of uncoated silica capillaries. The ionized silanol groups interact with the charged analytes and therefore after each separation any wall adsorpted materials must be cleansed off the capillaries. Even for coated capillaries (in CGE) periodic rinsing with the buffer, water and/or methanol is essential for best results.

Because of the small inner diameter, plugging causes problems in capillaries[4]. To avoid this potential problem, the dedicated capillaries can be stored with the ends immersed in buffer/distilled water, or capped with silicone rubber stoppers for up to about a week. To store capillaries for a longer time, dry storage is the best option. For drying, capillaries are first washed with 100 mM NaOH, rinsed with distilled water and then flushed with methanol, dried and stored.

To reuse a stored capillary, it should be given the same treatment as a new capillary. It is a good idea to rehydrate the lumenal surface of the dried capillary by rinsing first with distilled water and then with 100 mM NaOH and re-equilibrating with the buffer.

3.1.4 Coating:

A capillary is coated to eliminate or reduce the EOF. This also adds a degree of selectivity by either increasing or decreasing the adsorption of analytes to the walls. Coatings can take several forms, including simple addition of dynamic deactivation with buffer additives (that is, hydrophillic polymers or detergents) or covalent modification of the wall. Two fundamental approaches to limit solute adsorption are: a) permanent modification by covalently bonded or physically adhered phases; and b) dynamic deactivation using running buffer additives. An example of bonded or adhered phase would be neutral deactivation with polyacrylamide or polyethylene glycol, which eliminates EOF. This happens because of decreased effective wall charge and increased viscosity at the wall. These covalent modifications require little maintenance but repeated washings after every separation damages the coating.

Addition of modifiers to the running buffer is an alternative to the bonded or adhered phases. Surfactants are the most commonly used buffer additives. Surfactants provide the dynamic coating by either directly interacting with the solutes or by getting adsorbed to the capillary wall and thus, modifying EOF and limiting potential solute adsorption to the capillary wall. Dynamic coatings are not only more stable than adhered phases but also provide continuous regeneration of the coating by being in the running buffer.

Cellulose polymers produce a dynamic coating of the capillary, which helps to extend the column life. Table II lists some commercially available coated capillaries that can be used with many different kinds of instrumentation[3].

3.2 Injection systems:

In CGE, very small volumes of sample are loaded into the capillary in order to maintain high efficiency. A sample volume of 10-50 nL is common and volumes as small 100 pL have been reported. The sample plug length, which is a critical parameter, should, as a rule of thumb, be less than 1-2 % of the total length of the capillary.

3.2.1 Electroinjection or Electrokinetic injection:

A sample is loaded by replacing the injection-end reservoir with the sample solution and applying the voltage. The analyte enters the capillary by both migration and by the pumping action of the EOF. More mobile ions are loaded to a greater extent than those that are less mobile[5,6].

The quantities injected, Q (in gms or moles), can be calculated by the following equation:

$$Q = ((\mu_e + \mu_{EOF})\ V\pi r2Ct) / L$$

where:

μ_e = electophoretic mobility of the analyte

μ_{EOF} = EOF mobility

V = voltage

r = capillary radius

C = analyte concentration

t = time of the injection in seconds

L = total capillary length

Since it is possible to very accurately control both the voltage and the time, very good precision can be obtained by this technique. The presence of undetected ions like sodium or chloride can result in differences in voltage drop and quantity loaded (Figure 9). Due to this reason, electokinetic injection is generally not as reproducible but is very sensitive and produces sharp, well-resolved peaks. It is the best method of introducing samples into gel (cross-linked) filled capillaries. The difference in ionic strength between the sample and buffer produces a stacking effect that results in narrow sample zones (please refer to section 3.2.3 for explanation of stacking). This improves the sensitivity of the electrokinetic injection and results in improved resolution.

3.2.2 Hydrodynamic injection:

A sample can be loaded by application of pressure at the injection end of the capillary, vacuum at the exit end of the capillary, or by siphoning action achieved by keeping the injection reservoir a little higher than the exit reservoir for 20-30 seconds (Figure 10)[5,6]. To calculate the volume of sample loaded, the Hagen-Poiseiulle equation[1] can be used.

Volume = $\Delta P d^4 t / 128 \eta L$

where:

ΔP = pressure difference across the capillary

d = capillary inside diameter

t = time

η = buffer viscosity

L = total capillary length

Generally, the pressure used for loading ranges from 25-100 mbar and the time from 0.5-5 seconds. For siphoning injection, the pressure differential is given by:

$\Delta P = \rho g \Delta h$

where:

ρ = buffer density

g = gravitational constant

Δh = height differential of the reservoirs

Hydrodynamic injection tends to have lower precision than that found in electrokinetic injection.

3.2.3 Sample concentration:

On-capillary sample concentration, during or just after sample injection, enhances sensitivity. The conductivity differences between the sample zone and the running buffer gives rise to a phenomenon called "stacking" [3]. When voltage is applied, the greater field developed across the sample zone causes the ions to migrate faster till they reach the running buffer boundary and slow down. This process causes the concentration of the sample into a smaller zone.

3.3 Detection:

The small dimensions of capillaries present a serious challenge in detection. Although CE requires only a few nanoliter volume of the sample, it is not a "trace" analytical technique. Most of the time, concentrations of the analyte solution in the capillary or pre-concentration methods become necessary. Many different detectors have been used with CE but the most common is UV-Visible. The laser induced fluorescence (LIF) method is rapidly gaining popularity because of its high sensitivity. These two techniques are described in the following sections.

3.3.1 UV-Visible absorption:

For the majority of DNA applications, UV-absorbance at 254-260nm is normally used. The primary reason for its popularity is simply that the majority of compounds analyzed absorb somewhere in the UV region. Absorbance detection most often requires no chemical modification of the sample prior to analysis and is, therefore, an extremely convenient detection scheme. Sometimes, derivatization can be used to improve the detection of weak absorbers. Figure 11 illustrates a functional block diagram of a general CE/absorbance detector.

The detection limit is constrained by the small internal diameter of the capillary. The efficiency seen in case of CE, to some extent, is due to the on-capillary detection because the optical window is directly on the capillary so there is no zone broadening due to dead volume or analyte mixing after the separation[7]. For adsorptive detectors, the absorbance

of a solute, A, depends upon the path length, b, concentration, c, and molal adsorptivity, ε, as defined by Beer's law:

$$A = bc\varepsilon$$

3.3.2 Laser-Induced Fluorescence (LIF):

LIF is a far more sensitive detector than UV. A thousand fold increase in sensitivity over UV detection can be achieved easily. This high sensitivity is due to the fact that the analytical signal is measured against an extremely low background signal[3,8,9,10]. Because of this high sensitivity very narrow diameter capillaries can be used, thereby enhancing the resolution. Figure 12 shows a functional block diagram of an LIF detector for CE. In fluorescence, the analyte molecule absorbs a photon, and a fraction of the time, the excited molecule emits a photon upon returning to the ground state. A good fluorophore is a molecule that has high absorptivity at the exciting wavelength. The most important parameters, which determine the analytical utility of a fluorophore, are the absorptivity, fluorescence quantum yield, and photostability. The fluorescence quantum yield is the fraction of the excited molecules, which emit photons, therefore, having a high quantum yield, is also important. Photostability is generally reported in terms of the average number of excitation/emission cycles a molecule can undergo before being destroyed. The fluorescence detection systems have the best performance characteristics of any CE detection mode in terms of sensitivity, and limits of detection; however, many samples are not fluorescent. Thus, much of the published work deals with methods to convert the

samples into fluorescent forms by either derivatization or combining the fluorescence method with a less restrictive mode like immunoassay or indirect detection.

The most commonly used scheme seems to be the direct labeling of the analyte with a suitable fluorescent label (e.g. fluorescein isothiocyanate or FITC in short). Generally, DNA primers and probes are synthesized directly with the fluorescent labels attached to the 5′ end of the molecule. Laser light sources like He-Cd, He-Ne, and argon ion are generally used to fluoresce in the capillary. Table III lists several broad band sources and lasers with the wavelength range they provide for excitation. The laser light, which is monochromatic and very well collimated, is easily focused into the small capillaries. Stray light is rejected by collecting the emitted light at right angles to the excitation source, where rejection is done by spectral or spatial filtering.

4. Factors affecting CGE performance:

4.1 Gel :

The gel forms the sieving matrix for the separation of analytes. Several research groups have been investigating methods for the reliable production of reproducible, bubble-free gel-filled capillaries.

4.1.1 Chemical gels:

A chemical gel is a sieving matrix such as polyacrylamide (PA) which can be cross-linked and /or chemically bonded to the capillary wall. A PA gel can be produced from the polymerization of acrylamide monomers into long chains. Cross-linking of the long

chains occur by reaction with bifunctional compounds like bisacrylamide (N,N'-methelenebis(acrylamide)(Bis) [3,11,12,13]. Figure 13a shows the structure of the monomer and the gel and Figure 13b shows the covalent attachment of the gel to the capillary. The polymerization reaction is initiated by the addition of either ammonium persulfate or riboflavin. TEMED (N,N, N', N'-tetramethyl-ethylenediamine) is added to accelerate the polymerization process. Oxygen inhibits polymerization and so gel mixtures are usually degassed before use.

In the case of polyacrylamide gel, the pore-size can be increased by decreasing the acrylamide concentration or vice-versa. The total concentration of the monomer and of the cross-linking agent of a PA gel are expressed as[11]:

$$\%T = [\text{grams of acrylamide} + \text{grams of Bis}] \times 100/\text{Volume(ml) of solution}$$

$$\%C = \text{grams of Bis} \times 100 / [\text{gram of acrylamide} + \text{grams of Bis}]$$

Before adding the polymer solution, the fused-silica capillary is treated with a bifunctional reagent (e.g., 3-methacryloxy-propyl)-trimethoxy silane) which covalently binds the gel to the capillary.

In spite of chemical gels being highly effective in DNA separation, there are several drawbacks. Formation of air bubbles while preparing the gel, or filling the capillary, or during electrophoresis can lead to reduced resolution, a decrease in conduction and the breakdown of the gel structure. Other problems with these gels include degradation by hydrolysis, irreversible adsorption of higher molecular weight DNA's, and limited sample introduction.

4.1.2 Physical gels:

In this case, polymers used are not cross-linked and therefore can be easily pumped out of the capillary after each run. At a certain concentration known as the "entanglement threshold", the polymer strands begin to interact with each other resulting in a mesh like structure within the capillary[3,14]. For an effective separation the size of the mesh should approximate the size of an individual DNA fragment. The ideal polymer should be long enough to form a desirable mesh size, but short enough to keep viscosity at a minimum. Table IV lists some polymers used in CGE separations and their applications[3]. Several studies indicate that the addition of larger polymers improve the resolution of the high molecular weight DNA.

4.1.3 Cellulosic polymers:

Cellulose based polymers include methyl cellulose (MC), hydroxyethyl cellulose HEC), and hydroxypropylmethyl cellulose (HPMC). These polymers provide a much wider range of molecular weights, since they come from natural products[3]. These polymers also produce a dynamic coating of the capillary, thus helping extend the capillary's life. HEC is the most widely used cellulose polymer. The concentration of the polymer plays an important role in DNA separation. It is not the length but the molecular weight of the polymer that makes the matrix more or less viscous and as discussed before high viscosity increases the separation time and is thus undesirable. Figure 14 shows the separation of a DNA digest using a 1% solution of a low viscosity HEC. Because of the dynamic coating provided by cellulose polymers, the type of coating to be used with the

capillaries does not really matter significantly. But for best results, the capillaries should be rinsed with buffer, water, and/or methanol periodically.

4.2 Applied Voltage:

Even though applied potential of 10-20 kV is routinely used in the current practice, some research at higher voltage shows an increase in the migration rates, shorter analysis time, and well resolved peaks[15,16,17]. Some researchers have reported using a voltage as high as 1000 kV. In general, optimum resolution appears to occur at approximately 200V/cm. The biggest problem is Joule heat generation, which if not properly dissipated, cause density gradients in the gel, resulting in band broadening. Figure 15 shows the effect of voltage on separation efficiency and resolution.

4.3 Temperature:

Increasing temperature can have both positive and negative effects on the separation process in CGE. Increasing temperature mostly results in faster separations. High temperature decreases buffer viscosity which leads to an increase in solute electrophoretic mobility, and thus causes an increase in migration times[3,18]. Negative effects of increasing the temperature include sample denaturation, changes in buffer pH, and band broadening due to increase in diffusion, especially in case of hydrostatic injections. Most of the separations are done close to ambient temperature using water or air to keep the system thermally stable at that temperature.

4.4 Capillary size:

The most commonly used maximum capillary length is 100 cm. For quick and initial separation, shorter capillaries of approximately 20-50 cm work well but for significantly better resolution, an increased length of the capillary is necessary[3, 19].

Increasing the capillary i.d. increases the detection sensitivity due to an increase in the pathlength. This option is not usually desirable because a similar level of sensitivity can be achieved by on-line sample concentration phenomena during electrophoresis.

Increasing the i.d. also reduces the dissipation of Joule heat and hence band broadening results due to temperature gradients across the capillary.

4.5 Buffer/pH:

Selection of the right buffer for a separation in CGE plays a very critical role. Some of the most important qualities are: no interference with the analyte(s) of interest; low UV absorbency at the wavelength of interest; maintain solubility of the analyte(s); and maintenance of its buffering capacity through the separation. Table V lists some commonly used electrolytes in CE.

The buffer solution should be made in highly purified water using ultra-pure reagents. A buffer solution, stored at low temperature should always be first brought to the room temperature and then degassed before use.

Buffer additives can also be used to obtain increased or differential selectivity. Table VI lists some common buffer additives and their mode of action.

Choosing the buffer pH that is close to the pK of the solute mixture is ideal. With biological mixtures like DNA fragments, the pK_{avg} is typically close to neutral. The buffer pH can also be changed by use of additives, temperature, and ion depletion. Ion depletion generally occurs because of repetitive use of the same separation buffer. This depletion can result in poor resolution and reproducibility. A replenishment of the separation buffer solution after a couple of runs is desirable. Some researchers have seen that rinsing the capillary using a separate batch of the same buffer solution increases the number of runs before the loss of reproducibility sets-in.

Increasing the ionic strength of the separation buffer increases the thickness of the ionic double layer, and has the effect of decreasing the EOF and increasing the analysis time. This increase in analysis time increases the resolution in mixtures by decreasing non-specific analyte-analyte interactions. Increasing the ionic strength increases the current (this also means an increase in the production of Joule heat) at a constant voltage, and therefore, proper thermostating of the capillary becomes important[20]. Figure 16 shows the "Ohms Law Plot" that allows for easy determination of the "functional" buffer concentration and the maximum voltage that can be used with the particular buffer system. Linearity in a plot of observed current vs. voltage is an indication that the capillary temperature is being adequately maintained. Loss of linearity means that the thermostating capacity of the system has been exceeded. The heat generation of less than 1W/m (watts per meter) is desirable for optimum separation and should never exceed more than 5W/m. The figure shows a plot between the applied voltage and the current for 100 mM concentrations of each of the three buffers: borate, pH 8.3; phosphate, pH 2.5;

and CAPS, pH 11.0. At 20 kV the borate buffer has the lowest current of approximately 10μA while CAPS and phosphate have about 100 and 150 μA current respectively. The power associated with these buffers at 25kV is 0.58, 10.07 and 5.88 W/m for borate, phosphate and CAPS, respectively. Therefore borate buffer (100 mM) can be used at a high voltage of 30 kV (because it shows a linear relationship between current and the applied voltage) whereas the phosphate and the CAPS (both 100 mM) should not be used at voltages greater than 10 and 15 kV, respectively.

5. Modes of CE:

CE has various modes of separation that makes it a very versatile technique. These modes include capillary zone electrophoresis (CZE), micellar electrokinetic chromatography (MEKC), capillary isoelectric focusing (CIEF), capillary isotachophoresis (CITP), and capillary gel electrophoresis (CGE). Figure 17 shows the separation mechanism of each mode and Table VII describes the basis of separation.

5.1 CZE:

The separation occurs because solutes migrate at different velocities and in discrete zones. Neutral solutes coelute at the velocity of the EOF. Cations are attracted towards the cathode and their speed is increased by the EOF. Cations with higher charge to mass ratio elute first followed by lower charge to mass ratio cations. Anions on the other hand are attracted towards the anode but are pushed towards the cathode because of EOF. Anions with a small charge to mass ratio elute first. It is important to note that

charge/mass ratio can be modified by adjusting the pH of the buffer. CZE can be run with or without a wall coating. If effectively coated, the cations flow towards the cathode and anions flow toward the anode, while neutral molecules tend to remain at the injection site[5].

5.2 MEKC:

MEKC provides a mechanism to resolve neutral molecules. A detergent (e.g. sodium dodecyl sulfate, SDS) is used to generate micelles that are small droplets of oil with a highly polar surface (Figure 18). Neutral molecules partition in and out of the micelles based upon their hydrophobicities. For example, hydrophilic glucose will spend almost no time inside the micelle, while hydrophobic benzyl alcohol will spend a significant amount of time inside the micelle and will therefore be eluted later than glucose[5]. MEKC can be used for the separation of ionic and neutral compounds. It combines charge/mass ratio, hydrophobicity, and charge interactions at the surface of the micelles to affect the separation of analytes.

5.3 ITP:

Isotachophoresis is a useful technique for separating ionic materials only. It can separate either cations or anions in a single run. ITP requires the capillary wall to be coated to eliminate the EOF[4]. In ITP, the sample is sandwiched between leading and a terminating buffer (meaning solutions of different ionic mobility), which pulls the anions of the sample apart from each other (Figure 19). At the end of the run, the analytes separate with no space between them (See Figure 19).

5.4 IEF:

Isoelectric focusing has been used in slab as well as tube gel format for several years now[5]. The capillary wall is coated to eliminate EOF. A pH gradient is created inside the capillary with the cathode at the high pH side and anode at the low pH side of the gradient. Ampholytes are used to generate the pH gradient and are normally introduced with the sample. Zwitterionic analytes migrate to the position where their net charge is zero i.e., their isoelectric points (pI). The separated analytes can be removed by either positive pressure, vacuum or by changing the pH of the destination buffer. IEF can be easily used for zwitterionic compounds like proteins, amino acids, peptides, and several kinds of drugs.

6. DNA Basics:

6.1 DNA:

A DNA molecule consists of two strands that wrap around each other to resemble a twisted ladder, whose sides are made of sugar and phosphate molecules[21]. The two strands are connected to each other through hydrogen bonds between the bases on the strands. Each strand is a polymer consisting of nucleotides, which are composed of one sugar, one phosphate, and a nitrogenous base (Figure 20). The four different bases present in DNA are adenine (A), guanine (G), cytosine (C), and thymine (T). The first two are purine and the last two are pyrimidine bases.

A particular sequence of these bases specifies the exact genetic instructions required to create a particular organism with its own unique traits. The complete set of instructions

for making an organism is called a genome. The genome contains the blueprint for all cellular structures and activities for the lifetime of a cell or an organism. The size of the genome is stated in terms of the total number of base pairs (bps); the human genome contains roughly 3 billion bps. Figure 21 shows genome sizes of model organisms and humans.

6.2 Genes:

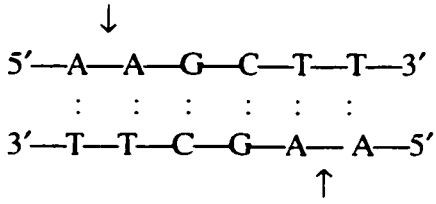
Each DNA molecule contains many genes; the physical and functional unity of heredity. Biochemically, a gene can be defined as that part of DNA (or in some cases RNA) that contains the information required to produce a functional biological product[21]. The product could be a protein or an RNA molecule. The human genome is estimated to consist of at least 100,000 genes.

6.3 Chromosomes:

"The three billion bps in the human genome are neatly organized into twenty four unique, physically separate microscopic units called chromosomes. All genes are arranged linearly along the chromosome. The nucleus of most human cells contains two sets of chromosomes, one set given by each parent. Each set has twenty two autosomes and an X or Y sex chromosome. Chromosomal DNA contains roughly one hundred and fifty million bases"[21].

6.4 Restriction nucleases:

Restriction nucleases (RN) are powerful tools for the manipulation of DNA. They are found in a wide range of bacterial species. The recognized sequences are almost always short (four to six bps, rarely more) and palindromic. Palindrome term is applied to those regions of DNA which have base sequences with twofold symmetry. They recognize certain DNA sequences and cut the DNA molecule at those specific sites[3]. For example, Hind III (a type of restriction enzyme) recognizes the sequence AAGCTT, and cuts the DNA between the two A's[22].



where the arrow symbol denotes the cutting site.

Some restriction enzymes cut the DNA infrequently, generating a small number of very large fragments (several thousand to a million bps). Most RNs cut DNA more frequently thus generating a large number of small fragments (less than a 100 to more than a 1000 bps). On average, RNs with 4, 6, and 8 base recognition sites will give fragments of 256, 4000, and 64000 bps, respectively. The distribution of fragment lengths is characteristic of the particular piece of DNA that was digested by the RN.

6.5 DNA Libraries:

Besides RNs, other molecular biology tools include cloning to introduce pieces of DNA into other organisms that are genetically engineered to be cloned easily. Because of

fragility, a chromosome is randomly broken into smaller pieces, which are then inserted into vectors for propagation. Vectors are DNA molecules originating from a virus, bacteria (plasmid) or the cell of a higher organism into which another DNA fragment of appropriate size can be integrated without loss of the vector's capacity for self-replication; vectors introduce foreign DNA into host cells, where it can be reproduced in large quantities[22]. The vectors can also be grown to make large quantities of the chromosomal fragments. The three most commonly used vectors are - plasmids, bacteriophages and cosmids. Plasmids (found in some species of bacteria) are circular DNA molecules that are found in the cytosol. Naturally occurring bacterial plasmids range in size from 5,000 to 400,000bps. With plasmids as vectors a DNA fragment of no more than 15,000 bps can be cloned. With bacteriophage λ (a type of bacterial virus) as a cloning vector DNA fragments up to 23,000 bps can be cloned. Cosmids are recombinant plasmids that combine useful features of both plasmids and bacteriophage λ. With cosmids DNA fragments up to 45,000 bps can be cloned[21].

It is a challenging job to recognize the randomly generated fragments of chromosomal DNA into an overlapping set that makes the whole chromosome. Before identifying a particular gene, a DNA library is constructed. A genomic DNA library contains DNA fragments derived from the genome of a particular organism. The information content of an organism in a library is represented by all the DNA fragments in it. After library construction, gene identification is done with the help of a probe. A probe is a

radioactively labeled DNA fragment that is complementary to the sequence of the gene of interest.

6.6 Sequencing Vector - M13mp18:

Even the cosmids are too large to be easily sequenced and thus are normally cut into smaller pieces, and these pieces are inserted into another vector - M13mp18 (a commonly used, genetically engineered bacteriophage). M13mp18 genome is a circular piece of DNA. Figure 22 shows the cloning steps more clearly[3].

To introduce a foreign DNA into the phage, it is first cut using a restriction nuclease (RN). The same RN is used to cut the phage vector. The cut fragments of foreign DNA and the vector are mixed and allowed to hybridize and then are permanently covalently joined with the help of a ligase enzyme. This vector containing the foreign DNA is now called a sequencing template. The sequencing vector generally contains ~2000 bases of foreign DNA and is used for the sequencing reaction.

6.7 Sanger's Chain termination reaction:

The dideoxy-chain termination reaction is an enzymatic method for DNA sequencing[22]. It involves controlled inhibition of enzymatic replication to produce DNA fragments of various sizes terminating at specific bases. A particular sequence of a single stranded DNA is copied using DNA polymerase I and is primed by a complementary fragment. A primer is a short oligonucleotide strand, complementary to the template with a 3′-hydroxyl group to which nucleotides can be added. This primer can be chemically

synthesized or obtained from restriction enzyme digestion. The four radioactively labeled deoxyribonucleoside triphosphates are incubated with a 2′, 3′ -dideoxy analog of one of them. This analog terminates the growth of a new chain because it lacks the 3′- hydroxy terminus to form the next phosphodiester bond and thus produces several fragments of different sizes with dideoxy at the 3′ end. Figure 23 shows this sequencing method in a very clear manner. Recently, research groups have come up with a slightly modified Sanger's technique for DNA sequencing. Instead of radioactive labeling, it uses a fluorescent tag attached to the primer- a differently colored one in each of the four chain terminating reaction mixtures[22]. The separated bands can be easily detected by their fluorescence and their color tells the identity of the base.

7. Analysis of DNA Fragments:

CGE is presently used as a tool of choice for the separation and sequencing of DNA fragments. The technique not only offers a fast, automated and efficient way to separate the oligonucleotides but also checks the purity. Synthetic oligonucleotides are extensively used as probes in the field of biochemistry and molecular biology for gene isolation and diagnostics, as primers for DNA sequencing, as linkers and adapters for cloning and gene alteration, site directed mutagenesis experiments and template amplification[23]. CGE is very helpful in purity analysis and sequence determination of these short synthetic oligonucleotides.

Before discussing the separation of DNA it would be helpful to understand the various mobility theories put forward by research groups to explain the most important concept in

capillary electrophoresis i.e., electrophoretic mobility. The mobility differences between neighboring bands in an electropherogram plays an important role in defining the resolution of the bands on a gel. Resolution varies with several different parameters and hence it is important to develop some kind of a theoretical model to explain the resolution of a DNA sequencing reaction using CGE. The most widely accepted theories or models are:

a) Ogston sieving model:

This theory treats polymer network as a molecular sieve[24]. It assumes that gel consists of random network of interconnected pores having an average pore size ξ and that the migrating solute behaves as an undeformable spherical particle of radius Rg. According to this model, smaller molecules migrate faster because more pores are available to them.

b) Reptation model:

This theory best describes the movement of long DNA molecules (where $Rg >> \xi$) through the gel network[24]. The long molecules snake through the gel "head first". The term reptation describes the alternating periods of stretching and relaxation that occur as the molecule moves through the small pores. This model works well in case of low electric field. In case of high electric field the DNA coil gets deformed and biased reptation model is used instead. The electric field influences the stretching periods of the DNA fragments, leading to a situation where the random walk becomes biased in the forward direction. When this stretching reaches a saturation level, all the size selectivity of the separation is lost.

Luckey and Smith[25] investigated the usefulness of these two models for the electric field strength ranging from 50 to 400 V/cm. They measured mobilities for ss DNA fragments ranging from 46 to 760 nucleotides in length. The gel solution used was 4% T and 5% C with 8.3 M urea. They found that both the above theories failed to accurately model their experimental results. They presented a modified Ogston theory which explains the size and field dependencies of the DNA mobility in polyacrylamide gels. Their theory explains the field dependence of mobility as due to a change in the effective size of the DNA. Due to the combined action of an applied electric field and the gel pores, the migrating DNA fragment gets stretched in the direction of motion. In other words, the DNA molecule distorts from a spherical shape to an ellipsoidal shape and thus decreases the effective radius of the migrating DNA (Figure 24). In the case of Luckey and Smith's experiment, an applied electric field helped DNA stretch in the direction of migration, but in the case of hydrodynamic measurements the ellipsoid may present its wider diameter to the gel pore because of a tumbling effect.

Comparison between ds DNA and ss DNA resolution:

Both ds DNA and ss DNA fragments can be separated using CGE with fairly good resolution. Schans et el[26] observed that ss DNA shows increased selectivity compared to separation of the same sample in its ds form. They found that the DNA separated in ss or denatured form gives better separation; meaning, a given separation can be achieved in a shorter period of time for ss DNA compared to separation of the same sample in its ds form. For DNA sequencing, only ss DNA is used.

Maniatis et el[27] compared the mobilities of ss and ds DNAs of same length and found that the ds DNA had a higher mobility than ss DNA. Figure 25 shows mobility of ss and ds DNA fragments at 200 V/cm and 400 V/cm respectively. The sample consists of 10 fragments of each, ds and ss DNA , ranging from 50 to 500 bps in 50 bps steps. 5′ end of both ss and ds DNA (only one strand) were labeled with fluorescein for LIF detection. This figure shows that the selectivity for ss DNA under denaturing conditions is higher than for ds DNA under ds or native conditions. Separation of ss DNA took less time (because of twice the voltage applied) but showed a wider migration window i.e., the difference between the migration times of the first and the last peak.

7.1 Separation of Homopolymeric oligonucleotides:

Figure 26 shows the separation of a polydeoxyadenylic acid mixture using CGE[28]. This polydeoxyadenylic acid mixture contains oligonucleotides ranging from 40-60 nucleotides. Baseline separation of 450 polydeoxyadenylic acids differing by only a single repeat unit has been reported by Cohen and Smisek[28]. The same team of researchers has also demonstrated baseline separation of picomole amounts of oligonucleotides analyzed in less than 8 minutes.

Figure 27 shows the separation of polydeoxythymidylic acids with 12-30 bases. The electropherogram reveals that the mixture is composed of three different component size groups, each one having a very distinct peak pattern[28]. The middle section, which shows a Gaussian distribution, consists of oligonucleotides containing 12 to 18 nucleotides. The

other two component groups show a relative decrease in the oligonucleotide amount with increasing nucleotide number. The pattern recognition allows an easy check of column performance and resolution.

7.2 Separation of Heterooligonucleotides:

Paulus and Ohms[23] examined the usefulness of gel-filled capillaries for the separation of heterooligonucleotides by chemically synthesizing a series of four oligonucleotides, two sets of complementary strands with 29 and 30 nucleotides each. The two strands of the two oligomers had the same base sequence. The only difference between the two oligomers was that the 30 base oligomer had one extra nucleotide. Thus, the 29 base oligomer could hybridize with its own complementary strand and also one of the 30 base oligomer strand, whereas hybridization with the second 30 base oligomer was not possible because of the same base sequence.

Figure 28 shows the electropherograms for the 29 and 30 base oligomers plotted together. The samples were injected right after their synthesis without any further purification. Each separation shows one major peak for the main component plus a number of smaller failure sequences with higher mobilities than the main product. Figure 28 clearly indicates the ability of CGE to separate the DNA fragments that differ by only one nucleotide.

The same experiment when performed after mixing the two oligonucleotide(29 and 30 base oligomer) mixtures still showed the baseline resolution (Figure 29). This experiment confirms that CGE can be successfully used for DNA sequencing.

Unlike homopolymeric separations, the heteropolymeric separations suffer from the hybridization problem. The heteropolymeric DNA fragments tend to show intramolecular base pairing, especially in the G and C (bases) rich regions. To inhibit hybridization, 7 M urea is used in the buffer and the capillary. In Figure 30, the separation shows the presence of three different species: two peaks between 18 and 19 min corresponding to the 29 and 30 base oligomer and an additional peak at 20 min. The presence of the third peak shows that in spite of adding urea, some of the complementary fragments hybridize. The appearance of the third peak depends on the time that the sample is left to hybridize. The separation shown in the above mentioned figure was obtained by injecting the sample 10 min after mixing. If the same sample is injected 70 min after mixing, the peak at 20 min becomes the major peak. After hybridization, the mixture consists of a double-stranded DNA fragment and a single-stranded, unhybridized oligonucleotide (the one present in excess).

7.3 Separation of long DNA fragments:

Large scale sequencing projects like the Human Genome Project (HGP) require techniques that can rapidly analyze the long DNA fragments. The sequencing process in CGE has been automated by using a fluorescence method of the Sanger dideoxy method but rapid analysis also requires an electrophoresis technique that is not only fast but can also resolve many bases at the same time[29].

The reading of long base sequences is important from the viewpoint of gene analysis because shorter groups must be overlapped many times to determine a long base

sequence. Several research groups have found that a high gel concentration can improve the separation of short sequences but not the long sequences and that a large migration distance (i.e. longer capillary) can improve the base resolution.

Nishikawa and Kambara[30]. showed that the resolution limit in a capillary with a 300 cm migration length, an electric field strength of 70 V/cm and a gel concentration of 4 % T, 3% C was 800 bases with resolution (Rs) = 0.5; however, this separation took 44 hours to finish. Resolution between adjacent peaks is defined as follows:

$$Rs = 1.18 \times d/2 \times w_{1/2}$$

where:

d = distance between two adjacent peaks

$w_{1/2}$ = DNA band half-width of a peak

If Rs is greater than 0.5, a valley appears between two adjacent peaks and if Rs is less than 0.5 then the two adjacent peaks are difficult to distinguish. Therefore, the base number at Rs = 0.5 can be used as an indication of the separation limit.

When the above separation was carried out in 200 cm long capillary at an electric field strength of 170 V/cm, a gel concentration of 3 % T, 5 % C, 680 bases with Rs = 0.5 were resolved in about 10 hours. As the base number increases by more than 600, the peak resolution decreases because of peak compression at large base numbers. The stretching of DNA fragments in the high electric fields causes this compression. Table VIII gives a comparison of base number (Nm) and migration time (Tm) of a fragment for both CGE and slab gel electrophoresis.

Figure 31 shows the separation of DNA fragments with base numbers from 60 to about 1000. The numbers above the peaks indicate the numbers of successive bases in a peak cluster.

Besides capillary length, gel composition and applied electric field can also be varied to influence the base resolution. Long base sequencing requires long migration distances in CGE, which implies a longer analysis time[30]. The high speed separation of DNA fragments using CGE can be achieved by increasing the electric field or by shortening the capillary length. CGE is generally operated at a high electric field of 100-300 V/cm. The problem with a high electric field is that it can reduce the resolution of long DNA fragments. Therefore, the challenge is to optimize the conditions (i.e. electric field and migration distance) for fast DNA separation. Figure 32 shows the separation of DNA fragments in a 4 % T, 5 % C polyacrylamide gel, capillary length of 30 cm, at electric field strength of 100 V/cm, 200 V/cm and 300 V/cm. The above mentioned figure shows that a high electric field strength results in rapid migration (short analysis time), but the resolution of long DNA fragments is lost. Some theoretical studies suggest that a low electric field (about 100 V/cm) and a large gel pore size will result in better separation of longer DNA fragments (upto 1000 bases).

Linear polyacrylamide (meaning no cross-linker) is increasingly becoming popular among many research groups because it is easy to replace after every separation[30]. Figure 33 shows the migration speed of a DNA fragment in a 9% T, 0% C matrix. Comparing this figure with Figure 32 shows that the speed of a DNA fragment in linear gels is 2-3 times

slower than that with a 4% T, 5% C gel. Also, the resolution obtained with the linear gel is lower than the resolution obtained with a 4% T, 5% C gel.

7.4 Comparison of slab gel electrophoresis (SGE) and CGE:

CGE and slab gel (open-bed) electrophoresis are two variations of the same basic separation technique. In general, CGE offers on-line or real time detection, easy quantitation, better resolution and efficiency. Earlier, researchers faced only one problem with CGE - only one sample can be analyzed at a given time. This problem has been solved by using several capillaries in parallel.

In CGE, run times are typically 30 min or less depending on the size of the oligonucleotides . Figure 34 compares the analysis times of the CGE and slab gel electrophoresis (SGE). In order to directly compare capillary gel and conventional slab gel techniques, identical samples were analyzed by the researchers on their own instrument and on a commercial fluorescent detection based machine (Applied Biosystems model 370A). Resolution was calculated using the IUPAC formula[31]:

$$R = \Delta T/4\sigma_T$$

Where ΔT is the difference in time of elution between two consecutive C peaks (differing by one nucleotide) and σ_T is the standard width of a single peak. Resolution values obtained using the above equation for the CGE ranged from 2.3 for the C33-C34 doublet, 1.8 for C156-157, and .8 for C215-216. SGE on the other hand gave 0.85, 0.6, and 0.4 resolution values for the same pairs.

The CGE separation was 3 times faster, taking 54 min for the peak labeled C100 to reach the detector, while the SGE required 150 min. The figure also shows that CGE offers better baseline resolution than SGE[32,33].

CGE is also better in terms of cost reduction, since it not only reuses the separation medium but also uses very small quantities of both the medium and the sample.

No additional time is required for detection and data reduction; sample preparation is the only manual step and accounts for 2 min per sample[32,33].

7.5 Problems with CGE:

Air bubble formation in the capillary, during polymerization, occurs about 20% of the time. Bubbles that are formed during electrophoresis, remain a serious problem. These latter bubbles form near the injection end of the capillary and often seriously affect resolution, especially of larger DNA fragments. In order to temporarily relieve the problem, a few centimeters of the capillary can be cut at the injection end after the sample has passed. Commercially produced quality-controlled gel filled capillaries can help reduce this air bubble problem to a certain extent.

At high voltages, electroendosmosis, caused by the presence of ionized silanol groups on the capillary's inner surface, can cause gels to exude out of the capillaries. This limits the voltages that can be used for enhanced performance and increased speed.

Formation of secondary structures in DNA fragments also affects the DNA mobility in capillary gels. Intramolecular base pairing at the 3′ end of single-stranded DNA molecules can cause an increase in electrophoretic mobility. In DNA sequencing

separations this can result in compressions, where several oligomers differing in length by one base comigrate. Konrad and Pentoney[33] tried incorporating formamide in the gel to reduce such compressions (Figure 35). They took a sequencing reaction containing three terminators, ddA, ddG, and ddC, in the approximate ratio of 4:2:1. Thus, with correct resolution, the largest peaks represented ddA terminated fragments, the intermediate peaks represented ddG, the smallest peaks ddC and the gaps ddT. The correct sequence through the region shown is CCCCGGGTACCGAGCTCGAA. The compressed region in the figure mentioned above is indicated by an arrow.

DNA sequencing gels typically contain 7 M urea as a denaturant, which reduces the stability of these secondary structures but does not completely remove them. Formamide, a stronger denaturant than urea, produces correct spacing for the peaks. Disadvantages of utilizing 35% formamide include increased separation times and reduced gel life. Figure 35c shows that the best separation is achieved by heating the capillary to 37°C during the run with much shorter separation times than the formamide gel.

8. Applications:

8.1 Purity control:

For small nucleotides (less than 10 bases) other modes of CE can be used for separation but for larger oligonucleotides CGE becomes the method of choice. Besides a purity check of synthetic nucleotides, CGE can also be used for their sequencing. CGE gives single base resolution for fragments up to 600 bps[2]. Generally PA is used as the sieving matrix because of availability and good resolving power. Highly viscous solutions of PA

and other physical gels can be polymerized inside the capillary. Besides factors discussed earlier that can affect resolution, one more important issue is "compression artifacts". In spite of thermal and/or chemical denaturants, short synthetic oligomers form some secondary structures (Figure 36) that affect their mobility in the capillary during separation.

8.2 Anti-sense DNA:

CGE is often used to determine the quality of chemically synthesized DNA analogs[34], which are currently being studied as antisense therapeutics. Antisense therapeutics are synthetic oligonucleotides whose sequence of bases is complementary to a target sequence on either a mRNA, which encodes for disease causing proteins, or the double-stranded DNA from which the mRNA was transcribed. This antisense oligonucleotide can hydrogen bond and thus inactivate the genetic message. Some examples of these therapeutics are, phosphorothioates, methyl phosphonates, and peptide nucleic acids (PNA). CGE provides single base resolution, easy on-line detection, and ease of sample injection into the capillary.

8.3 PCR product analysis:

The polymerase chain reaction (PCR) is a method of enzymatically amplifying a desired DNA sequence of any origin (virus, bacteria, human, or plant) millions of times in a matter of hours. PCR analysis provides a highly specific, automated capability of amplifying very small amounts of any sample. Purification of PCR products before their analysis is required to remove the salts contained in the samples. These salts can cause

band broadening due to the conductivity deference between the sample zone and the surrounding buffer. Since CGE provides the ability to rapidly quantitate the products, it has become a technique of choice for PCR product analysis[3].

Applications of PCR analysis include genetic disease diagnostics, forensic science, and evolutionary biology to name a few. PCR treated blood samples, for example, can be analyzed using CGE, to screen for HIV or for human identification (in civil and criminal cases).

8.4 Large DNA separation:

CGE has been successfully used for large DNA fragments, meaning a double-stranded DNA molecule with more than 2000 bps. Various researchers have demonstrated the use of high voltage, steady fields in dilute HEC solutions for the separation of large DNA fragments. Normally fast separation of DNA fragments up to 700 bases requires a low electrical field rather than long migration time. A sieving matrix having large gel pores can increase separation of large DNA fragments[3].

8.5 Human Genome Project (HGP):

CGE has seen a very rapid growth and development in the last few years because of the HGP. It has been extensively used for large scale DNA separation and sequencing. Recent use of multiplexed capillary DNA sequencers has increased the speed and efficiency of DNA separation. Various technological advancements in the development of CGE such as easily replaceable, non-crosslinked matrix (gel), fluorescent detection methods and labeling techniques etc., have made CGE a very attractive separation

technique for DNA fragments for large scale sequencing projects like the HGP. A next generation system under construction will operate 96 capillaries simultaneously, which will eventually be expanded to 864 capillaries. An 864 capillary instrument will be able to separate DNA fragments up to half a million bases in one and a half-hours.

HGP is an international research program to construct detailed genetic and physical maps of the human genome and several other model organisms used in research laboratories. Mapping includes the determination of the complete sequence of human DNA, localization of the estimated 50,000-100,000 genes within the human and other organisms genomes[35].

The HGP began in the mid-1980's and is projected to be completed by sometime in 2005. The Department of Energy (DOE) and the National Institutes of Health (NIH) are the two main research agencies within the US government responsible for developing and planning the project. They jointly published their research plan in, "Understanding Our Genetic Inheritance: The US Human Genome Project. The First Five Years FY 1991-1995."[35]

The most important goal[35,36,37] of the HGP is to make a descriptive map of each chromosome at increasingly finer resolutions. Mapping involves fragmenting the chromosome into several little pieces that are propagated and characterized, and then ordered (mapped) to correspond to their respective locations on the chromosomes. After mapping, the base sequence of each DNA fragment is determined. All this information about the DNA sequence and the location of genes in the DNA sequence will be used to develop the tools for the study of human biology and medicine. The location of

approximately 2300 genes has already been determined, and the effort to generate a high-resolution map of the genome is also proceeding according to the plan.

The detailed information about the nucleotide sequence of DNA will help in understanding the structure, organization, and function of DNA in chromosomes. Genome maps of the model organisms will help in the understanding of the function of the genes in health and disease[37,38,39].

Genes involved in various genetic diseases like cystic fibrosis, myotonic dystrophy, Duchenne muscular dystrophy, neurofibromatosis, and retinoblastoma have already been identified. More research will uncover the mechanisms for diseases caused by several genes or by a single gene interacting with environmental factors. Emphasis would shift from treatment of the sick to a prevention -based approach. Individuals predisposed to certain diseases would have the option of defective gene replacement through gene therapy.

Technologies, databases, and biological resources developed in genome research will have an enormous impact on a wide variety of biotechnology-related industries. The data that will accumulate from genome research could present social , ethical and legal dilemmas. To handle those kinds of issues the HGP is funding conferences and research projects to identify and consider relevant issues, and also to promote public awareness about these issues.

9. Conclusions:

CGE produces very high resolution separation of DNA molecules; even fragments that differ in size by only a single nucleotide can be resolved. Almost all steps in separation and sequencing are now automated. Still, these first generation gel-based technologies that have been used for the last few years have only been successful on sequencing small regions of interest in the human genome.

The need to develop a sequencing technology that is highly sensitive, accurate, and economical is even greater now because of the on-going HGP. The HGP plan is to develop a technique that can sequence 100,000 bases/day at a cost of less than 50 cents/base. The second generation gel-based sequencing technologies will use high voltage capillary and ultra-thin electrophoresis to increase fragment separation rates. CGE has come a long way from being a fancy analytical technique to a routinely used separation and sequencing technique. CGE can provide a higher throughput by use of multiplexed arrays of 20-100 capillaries simultaneously undergoing electrophoresis. With the recent advances in the field of detection techniques and various types of polymer networks, CGE will soon become the tool of choice in the fields of diagnostics, biomedicine, forensic science, and biotechnology.

10. References

1. D.N. Heiger, *High Performance Capillary Electrophoresis*, Hewlett-Packard Company, France, (1992)
2. C. A. Monnig, R. T. Kennedy, *Anal. Chem.*, 66, 280R-314R, (1994)
3. J. P. Landers , *Handbook of Capillary Electrophoresis*, CRC Press, 2nd Ed., Chap 1, 11, 19, (1996)
4. R. F. Boyer, *Modern Experimental Biochemistry*, The Benjamin/Cummings Publishing Company, 2nd Ed.
5. J.D. Olechno, J.M.Y. Tso, J. Thayer , A. Wainright, *Am. Lab.*, 22 [18], 30-37, (1990)
6. M.J. Gordon, X. C. Huang, S. L. Pentoney Jr., R. N. Zare, *Science*, 242, 224-228 (1988)
7. X.C. Huang, S. G. Stuart, P. F. Bente III, T. M. Brennan, *J. of Chromatogr.*, 600, 289-295, (1992)
8. T. Nishikawa, H. Kambara, *Electrophoresis*, 13, 495-499, (1992)
9. A. Belenky, D. L. Smisek, A. S. Cohen, *J. of Chromatogr.*, 700, 137-149, (1995)
10. D. Y. Chen, H. P. Swerdlow, H. R. Harke, H. Z. Zhang, N. J. Dovichi, *J. of Chromatogr.*,559, 237-246, (1991)
11. Y. Baba, M. Tsuhako, *Trends in Analytical Chemisty*, vol. 11.no. 8, 281-287 (1992)
12. H. R. Harke, S. Bay, J. Z. Zhang, M. J. Rocheleau, N. J. Dovichi, *J. of Chromatogr.* 608, 143-150, (1992)
13. D. Figeys, N. J. Dovichi, *J. of. Chromatogr.*, 645, 311-317, (1993)
14. Y. Cordier, O. Roch, P. Cordier, R. Bischoff, *J. of Chromatogr.*, 680, 479-489, (1994)
15. P. Sun, R. A. Hartwick, *J. of Liquid Chromatogr.*, 17(9), 181-1875, (1994)
16. J. A. Luckey, L. M. Smith, *Electorphoresis*, 14, 492-501, (1993)
17. G. W. Slater, P. Mayer, G. Drouin, *Electrophoresis*, 14, 961-966, (1993)
18. A. Guttman, N. Cooke, *J. of Chromatogr.*, 559, 285-294, (1991)
19. R. L. Karger, Nature, Vol. 339, 641-642, (1989)
20. A. Guttman, E. Szoko, *J. of Chromatogr.*, 744, 321-324, (1996)
21. http://www.ornl.gov/TechResources
22. L. Stryer, Biochemistry, Third Edition, Chap 6 (1988)
23. A. Paulus, J. I. Ohms, *J. of Chromatogr.*, 507, 113-123, (1990)
24. A.E. Barron, H.W. Blanch, D.S. Soane, *Electorphoresis*, 15, 597-615, 1994
25. J.A Luckey, L.M. Smith, *Electrophoresis*, 14, 492-501, 1993
26. M.J. van der Schans, A.W.H.M. Kuypers, A.D. Kloosterman, H.J.T. Janssen, *J. of Chromatography A, 772, 255-264, 1997*
27. T. Maniatis, A. Jeffrey, H. van de Sande, Biochemistry, 14, 3787, 1975
28. A. S. Cohen, D. L. Smisek, *Trends in Analytical Chemistry*, vol. 12, no. 5, 195-202, (1993)
29. M. Kamahori, H. Kambara, *Electrophoresis*, 17, 1476-1484, (1996)
30. T. Nishikawa, H. Kambara, *Electrophoresis*, 15, 215-220, (1994)

31. H. Swerdlow, R. Gesteland, *Nucleic Acids Research*, Vol. 18, No. 6, 1415-1419 (1990)
32. K. D. Konrad, S. L. Pentoney, Jr., *Electrophoresis*, 14, 502-508, (1993)
33. J. A. Luckey, H. Drossman, A. J. Kostichka, D. A. Mead, J. D'Cunha, T. B. Norris, L. M. Smith, *Nucleic Acids Research*, Vol. 18, No. 15, 4417-4421 (1990)
34. L. A. DeDionisio, D. H. Lloyd, *J. of Chromatogr.*, 735, 191-208, (1996)
35. http://www.ornl.gov/TechResources/Human_Genome/project/project.html
36. http://www.ornl.gov/TechResources/Human_Genome/project/hgp.html
37. http://www.nhgri.nih.gov/HGP
38. http://www.gene.com/ae/AB/IE/What_Can_We_Expect.html
39. http://www.gene.com/ae/AB/IE/Intro_The_Human_Genome.html

11. List Of Abbreviations

CE	capillary electrophoresis
CGE	capillary gel electrophoresis
DNA	deoxyribonucleic acid
PCR	polymerase chain reaction
HGP	human genome project
ss DNA	single-stranded DNA
ds DNA	double-stranded DNA
μm	micron
EOF	electroosmotic flow
CZE	capillary zone electrophoresis
MEKC	micellar electrokinetic chromatography
ITP	isotachophoresis
IEF	isoelectric focusing
LIF	Laser induced fluorescence
Bps	base pairs
RN	restriction nuclease
FITC	fluorescein isothiocyanate
DOE	department of energy
NIH	national institutes of health

Electrophoresis is performed in narrow-bore (25- to 75-μm id), fused silica capillaries

High voltages (10 to 30 kV) and high electric fields (100 to 500 V/cm) are applied across the capillary

High resistance of the capillary limits current generation and internal heating

High efficiency ($N > 10^5$ to 10^6) and short analysis time

Detection performed on-capillary (no external detection cell)

Small sample volume required (1 to 50 nl injected)

Numerous modes to vary selectivity and wide application range

Operates in aqueous media

Simple methods development

Automated instrumentation

Figure 1. Characteristics of CE[1]

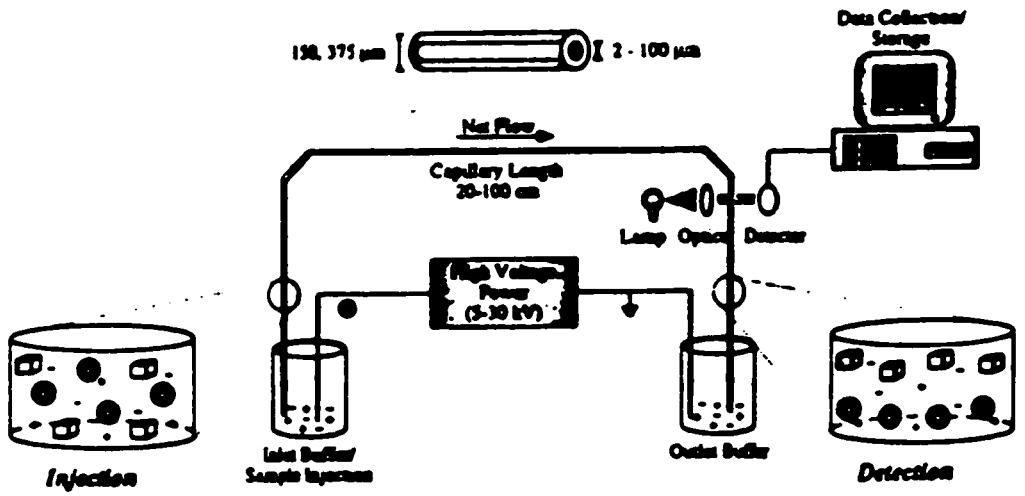


Figure 2. General schematic of a CE instrument[3]

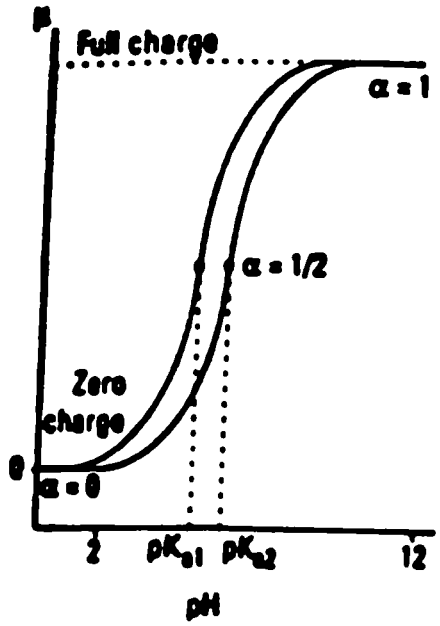


Figure 3. Mobility of 2 weak acids as a function of pH[1]

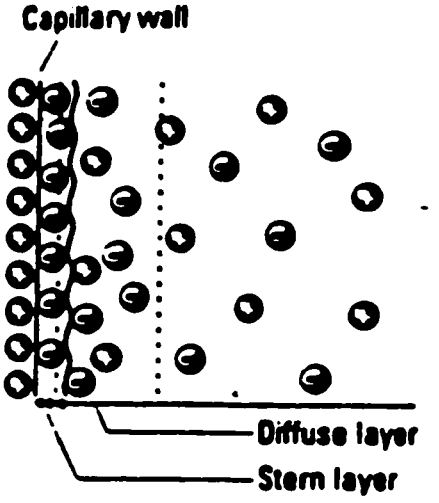


Figure 4. Representation of the double layer at the capillary wall[1]

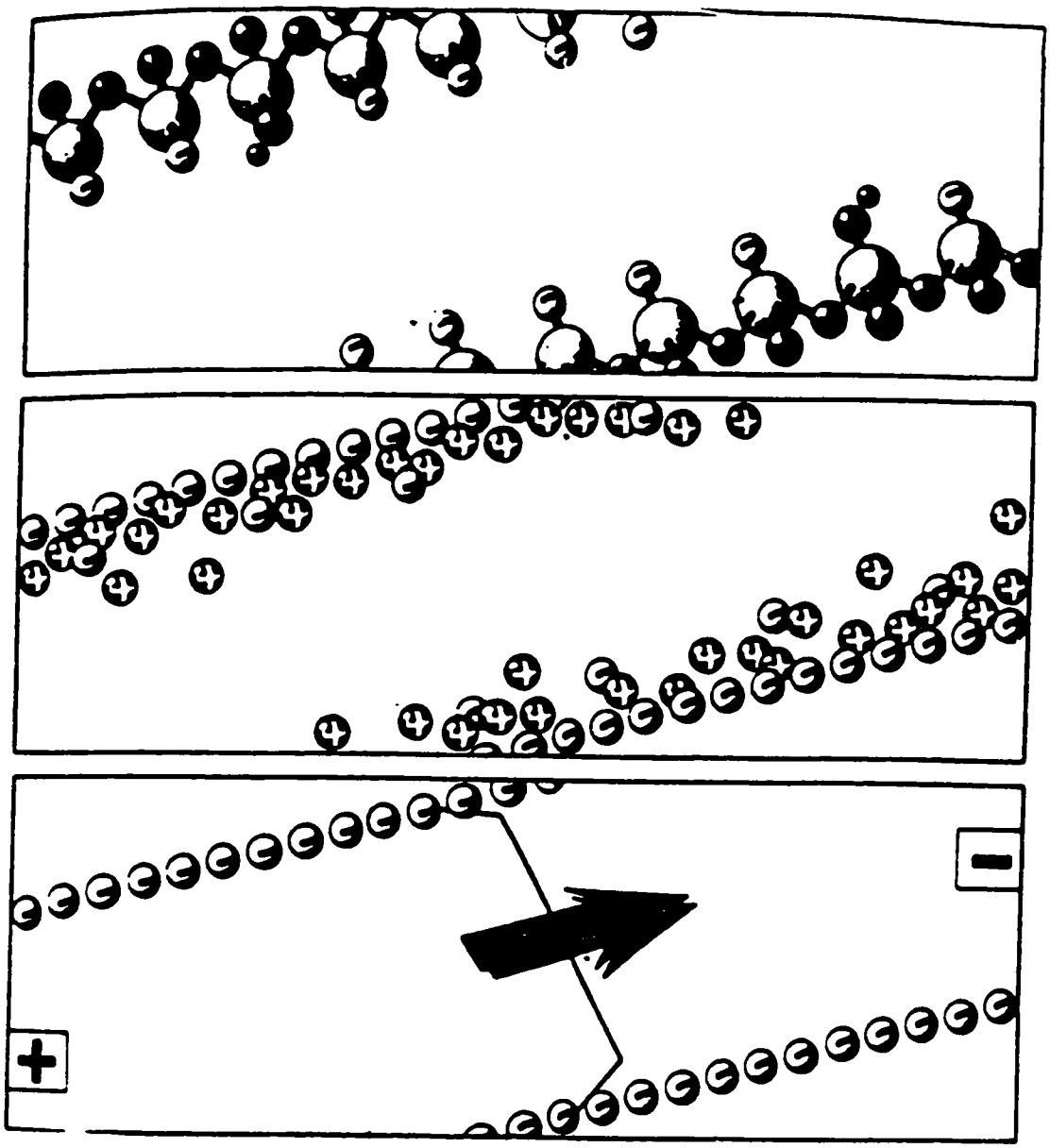


Figure 5. Development of the electroosmotic flow[1]

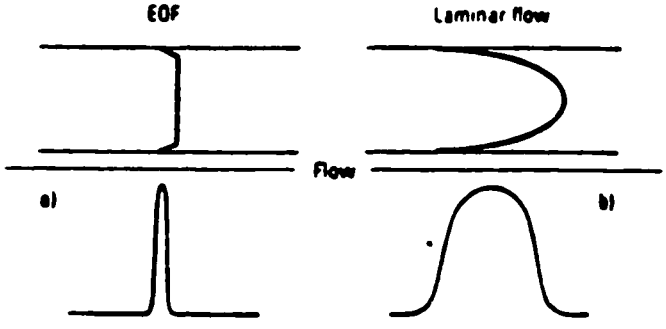


Figure 6. Flow profile and corresponding solute zone[1]

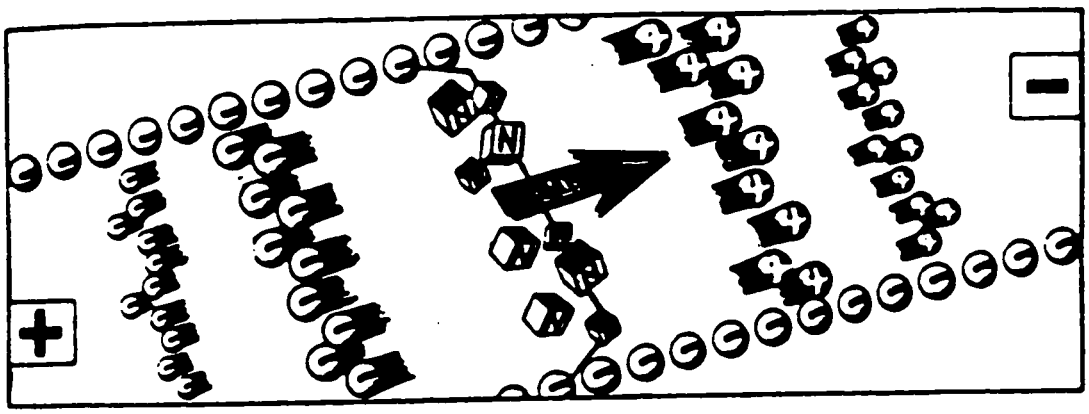


Figure 7. Differential solute migration superimposed on EOF in CZE

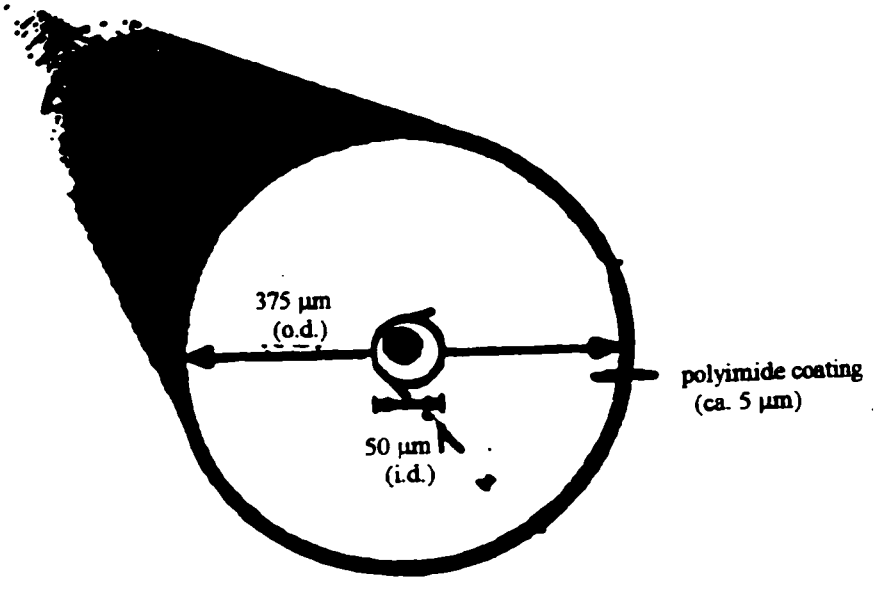


Figure 8. End view of a capillary with a 375 µm o.d. and a 50 µm i.d. [3]

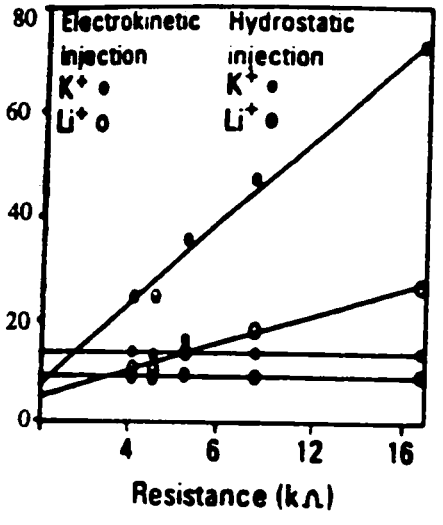


Figure 9. Quantity of sample loaded as a function of sample resistance for hydrodynamic and electrokinetic injection[1]

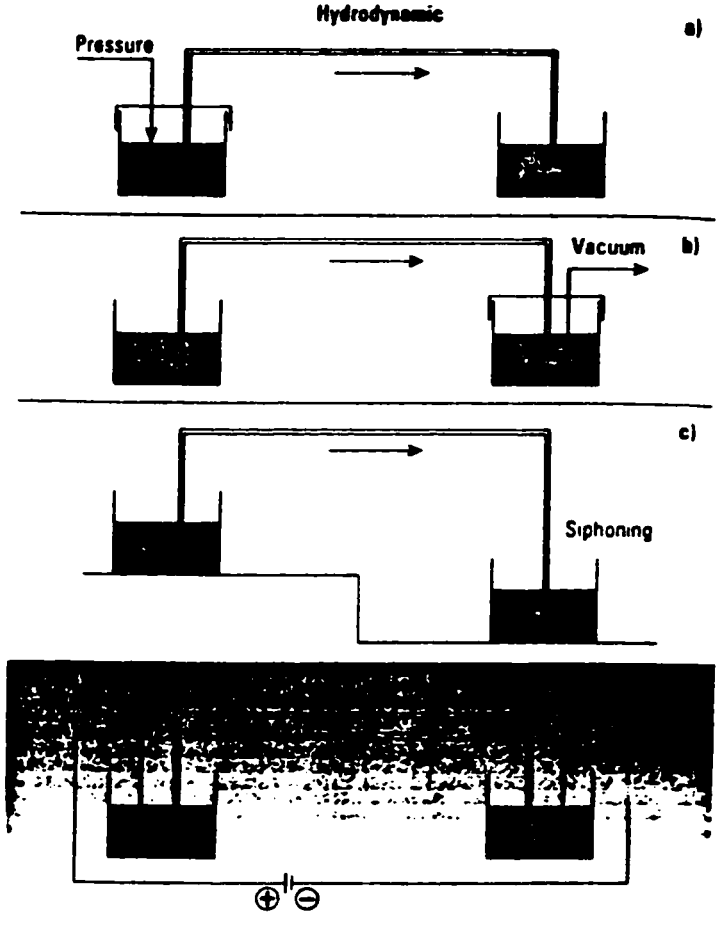


Figure 10. Methods of sample injection[1]

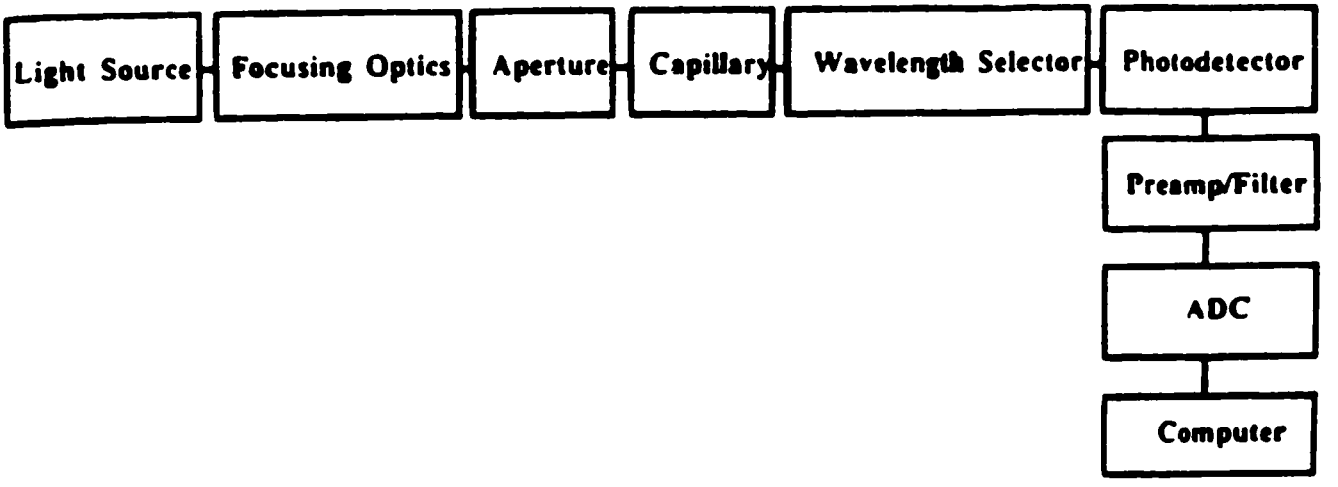


Figure 11. Functional block diagram of an UV-Vis absorbance detector for CE[3]

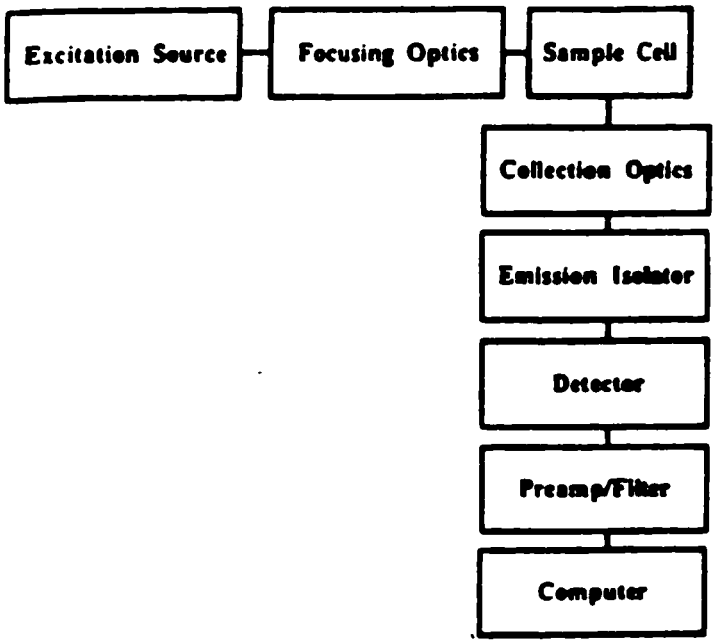


Figure 12. Functional block diagram of an LIF detector for CE[3]

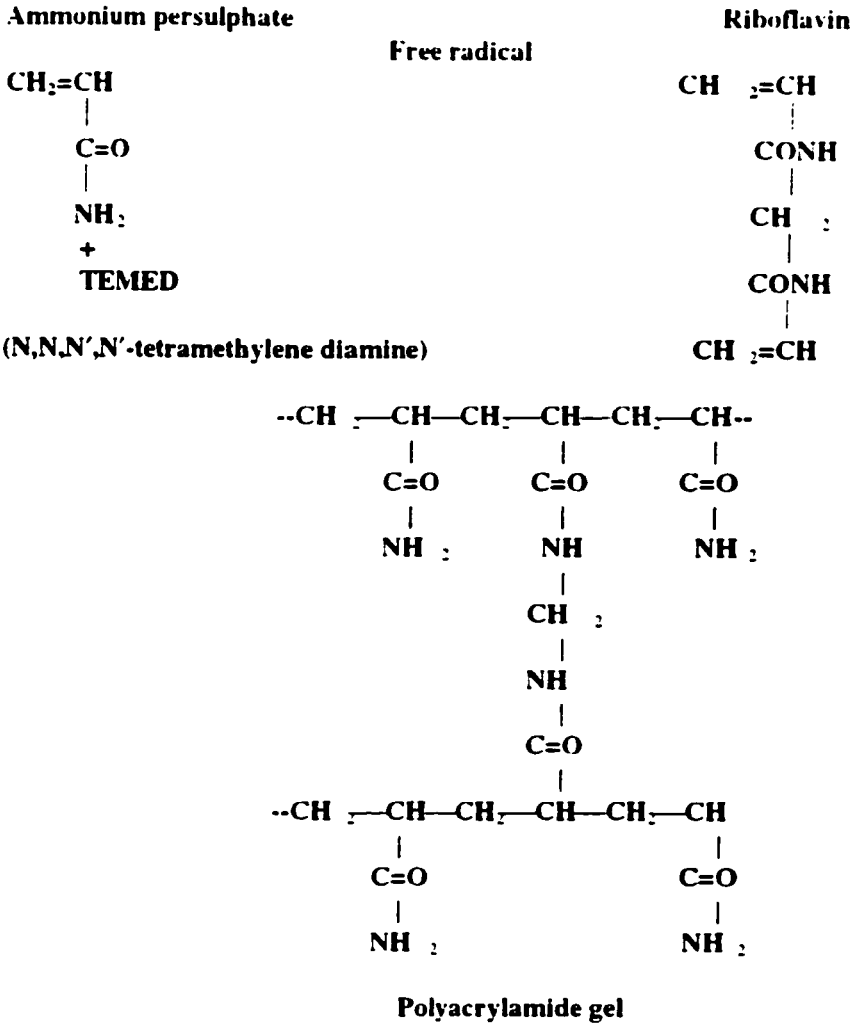


Figure 13a. Structure of monomer and the gel[11]

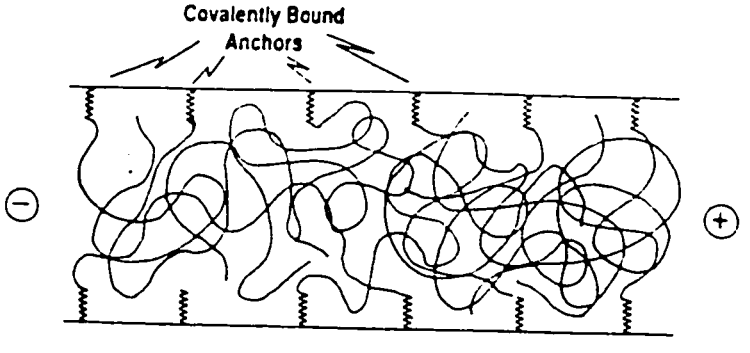


Figure 13b. Covalent gel attachment in CGE[11]

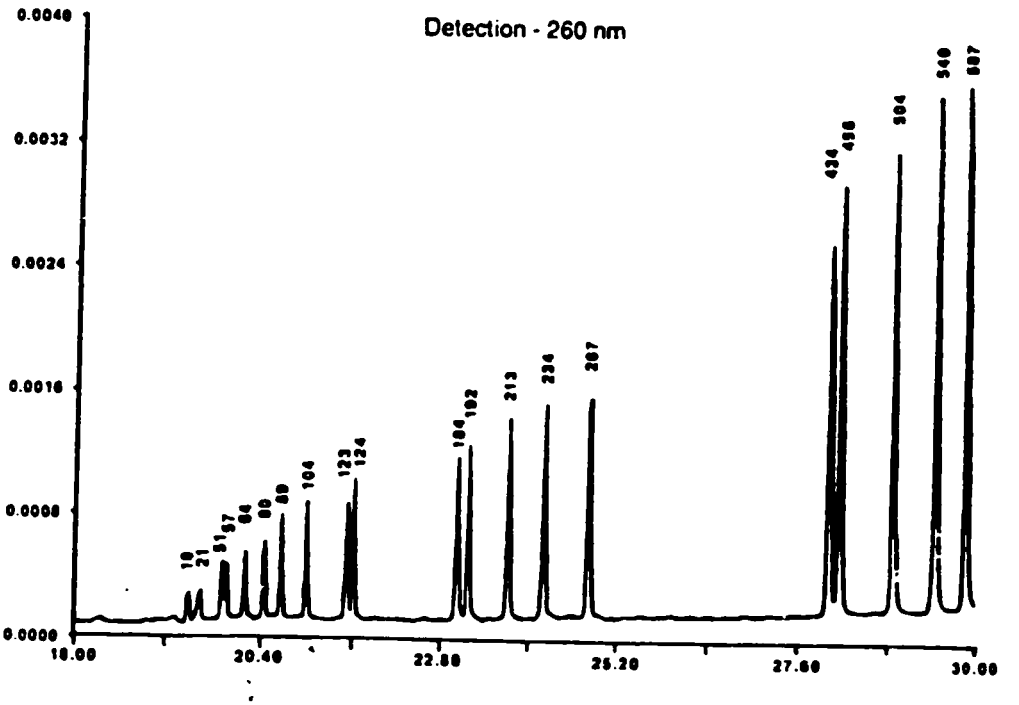


Figure 14. Separation of pBR 322 HaeIII digest using a cellulosic polymer[3]
Conditions: 60 cm, 100 μm i.d., DB-17 capillary; 2 seconds electrokinetic injection of sample at 10 kV; 38 μA constant current 100 m*M* TBE buffer and 6 μ*M* ethidium bromide with 1% HEC; UV detection at 260 nm.

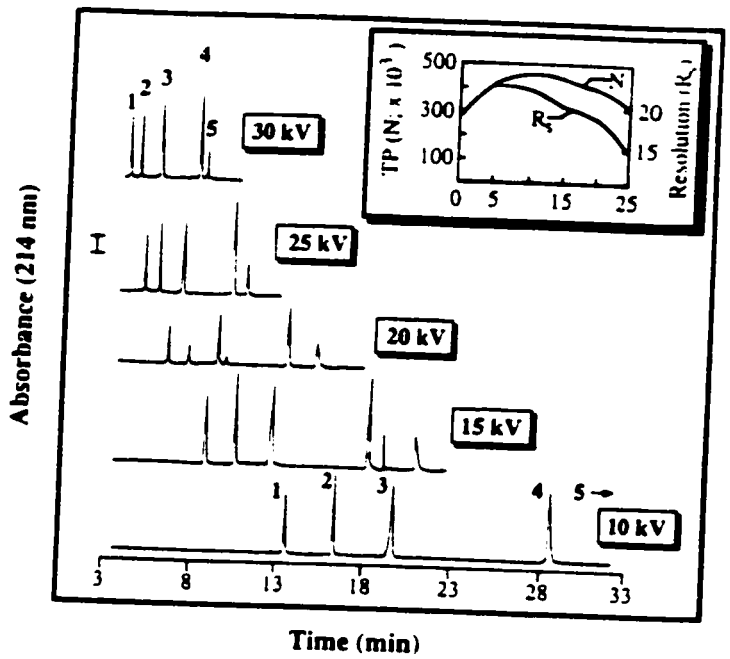


Figure 15. Effect of voltage on separation efficiency and resolution[3]

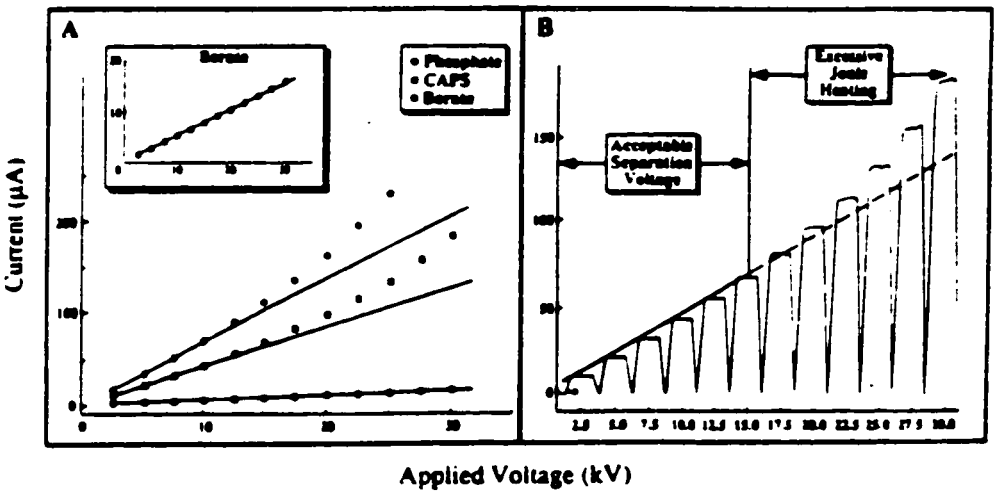


Figure 16. Ohm's Law Plot[3]

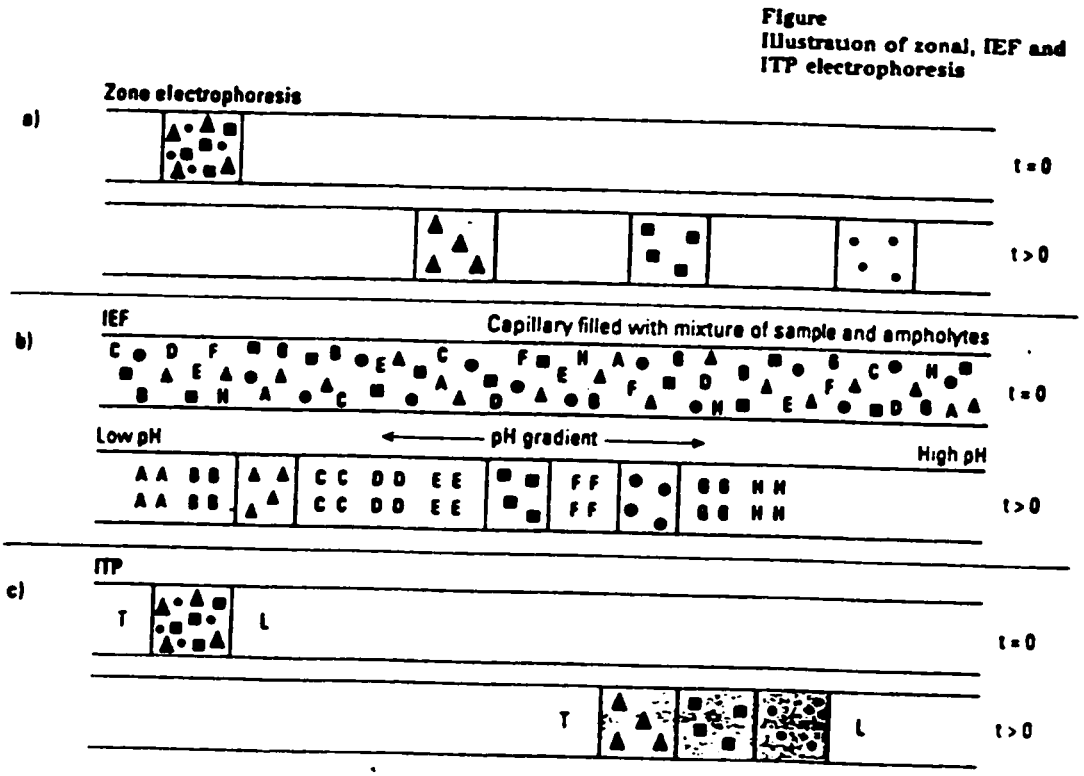


Figure 17. Illustration of CZE, IEF and ITP electrophoresis

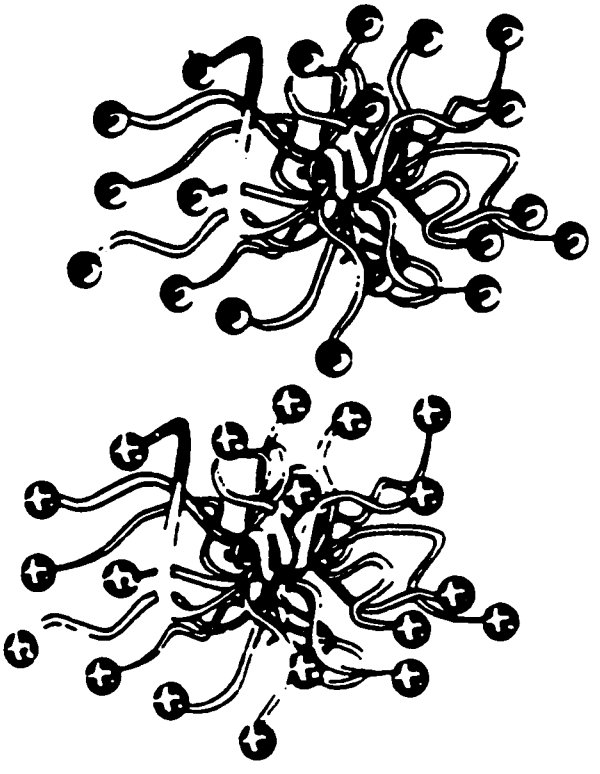


Figure 18. Schematics of cationic and anionic micelles[1]

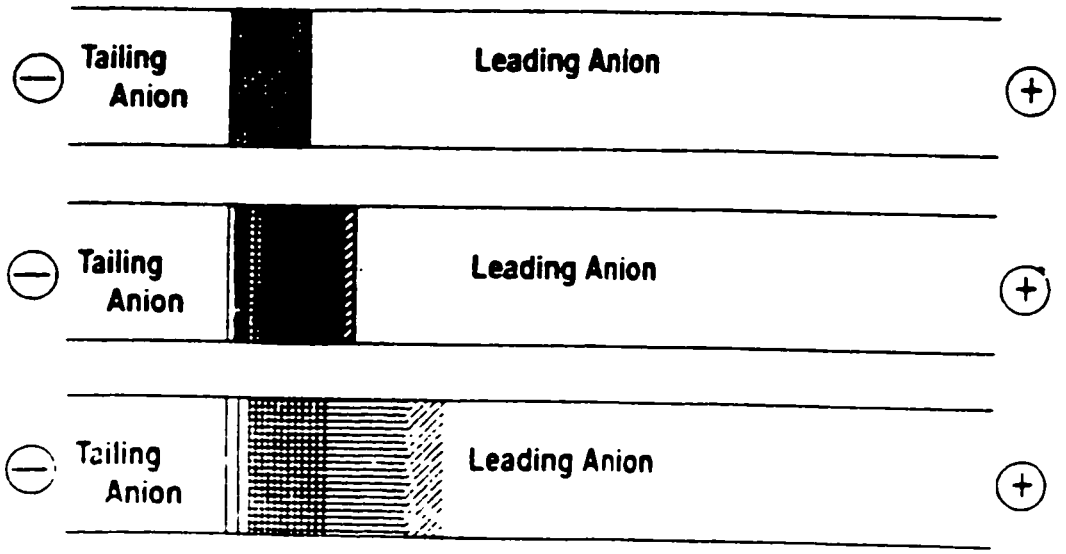


Figure 19. Separation by ITP[5]

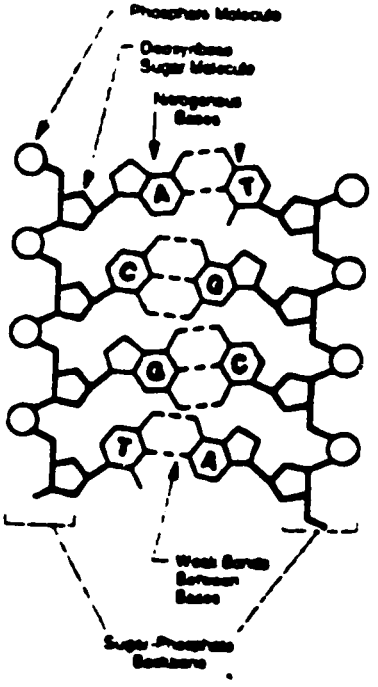


Figure 20. DNA structure[21]

Comparative Sequence Sizes	(Bases)
(yeast chromosome 3)	350 Thousand
Escherichia coli (bacterium) genome	4.6 Million
Largest yeast chromosome now mapped	5.8 Million
Entire yeast genome	15 Million
Smallest human chromosome (Y)	50 Million
Largest human chromosome (1)	250 Million
Entire human genome	3 Billion

Figure 21. Comparison of largest known DNA sequence with approximate chromosome and genome sizes of model organisms and humans[35]

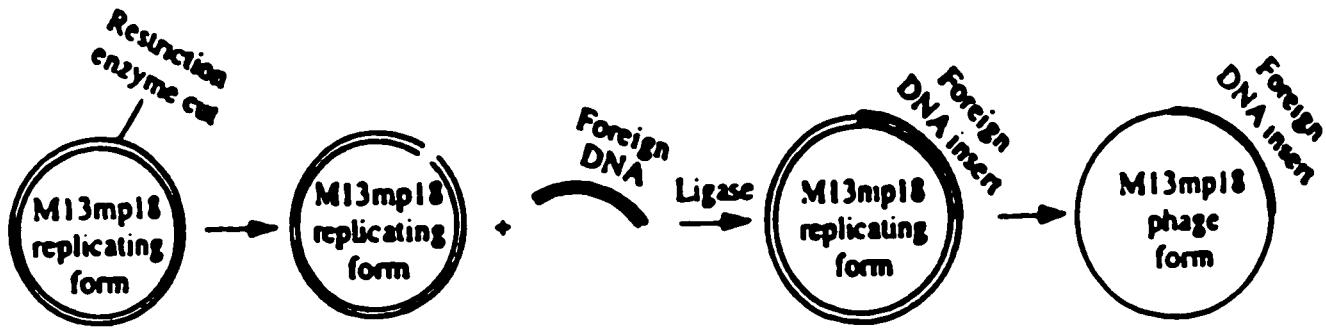


Figure 22. Cloning of foreign DNA into M13mp18. [3]

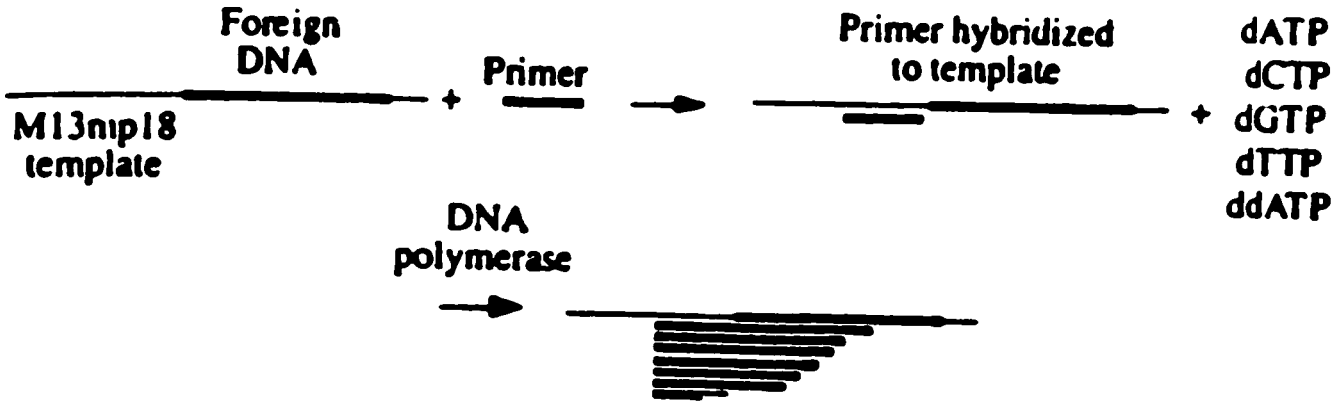


Figure 23. Sanger's di-deoxy chain termination sequencing reaction[3]

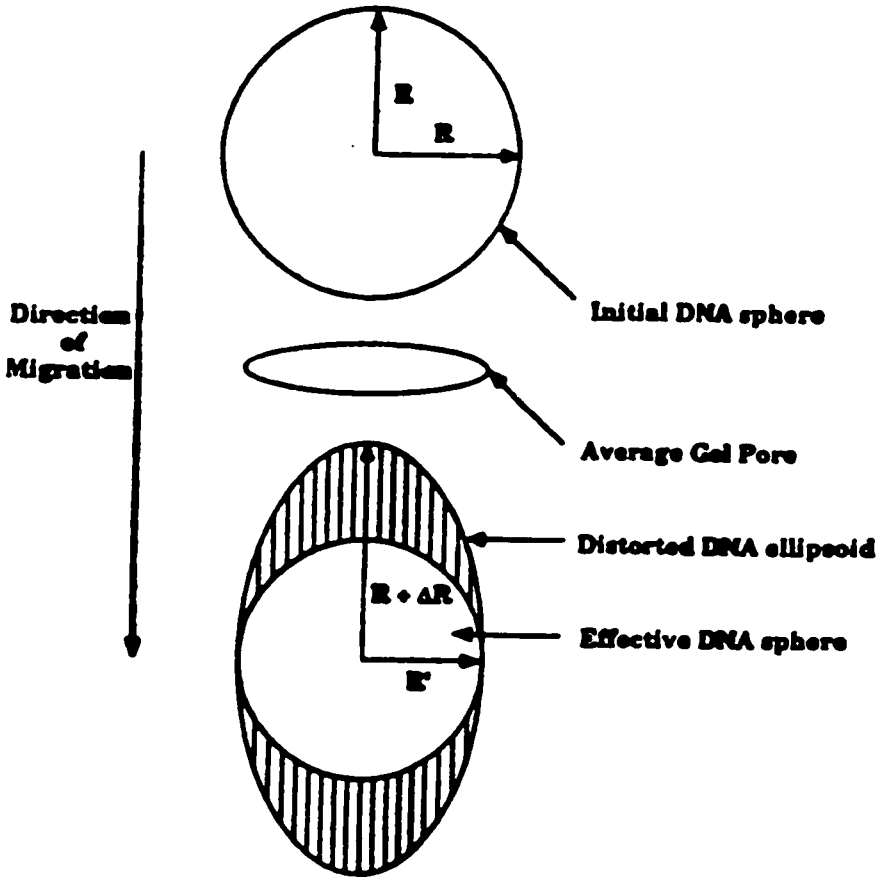


Figure 24. Schematic of the presumed deformation of DNA molecules that occurs in the presence of an eiectric field in polyacrylamide gels. An initial random coil of radius R is pulled through a relatively rigid gel pore by the action of the electric field. This causes the DNA to elongate and form a spherical ellipsoid with minor radius R'. It is assumed that this distorted molecule is functionally equivalent to a DNA sphere of radius R' as shown[25].

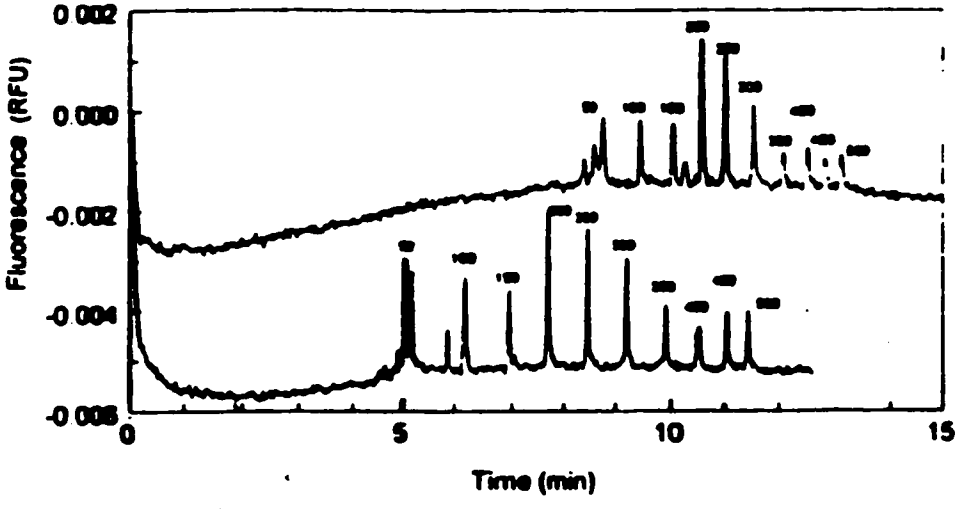


Figure 25. Electropherogram of DNA 50-500 sizer from Pharmacia. Upper electropherogram: native conditions dsDNA, voltage = 7.4 kV. Lower electropherogram: denaturing conditions ssDNA, voltage = 14.8 kV, capillary: I.D. = 75 μm, *l* = 30/37 cm, pressure injection of 5 s, temperature = 20°C[26]

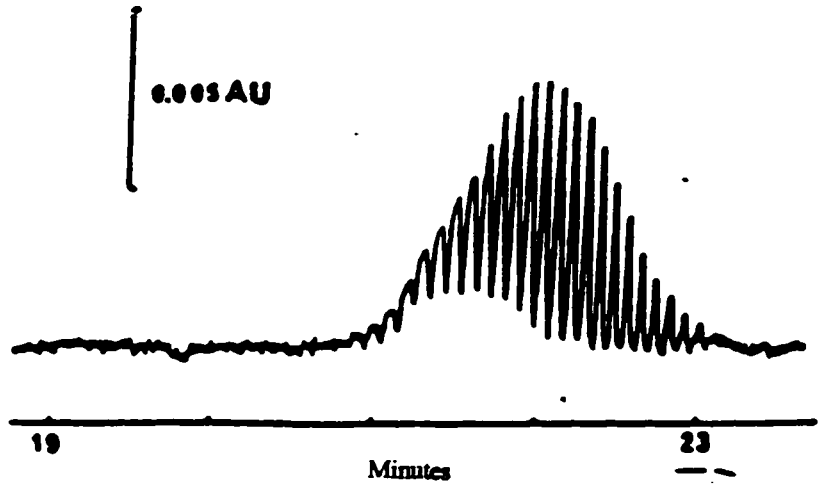


Figure 26. Micropreparative CGE separation of a polydeoxyadenylic acid test mixture, $p(dA)_{40-60}$[28]

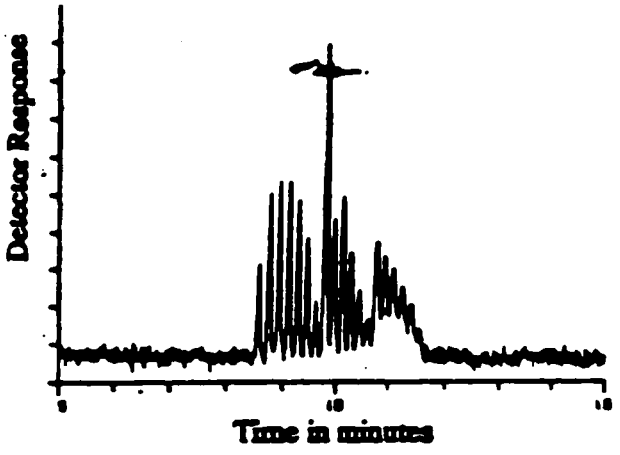


Figure 27. Separation of deoxypolythymidylic acid pd(T) 12-30[23]

Separation conditions: capillary 75μm I.D., 360 μm O.D., length 40 cm, effective length 20 cm; buffer 50 m*M* Tris, 50 m*M* boric acid, 7 *M* Urea; column 7.5% T, 3.3 % C; field 375 V/cm; current 5.6 μA; injection 4000 Vs; detection 260 nm UV on-column

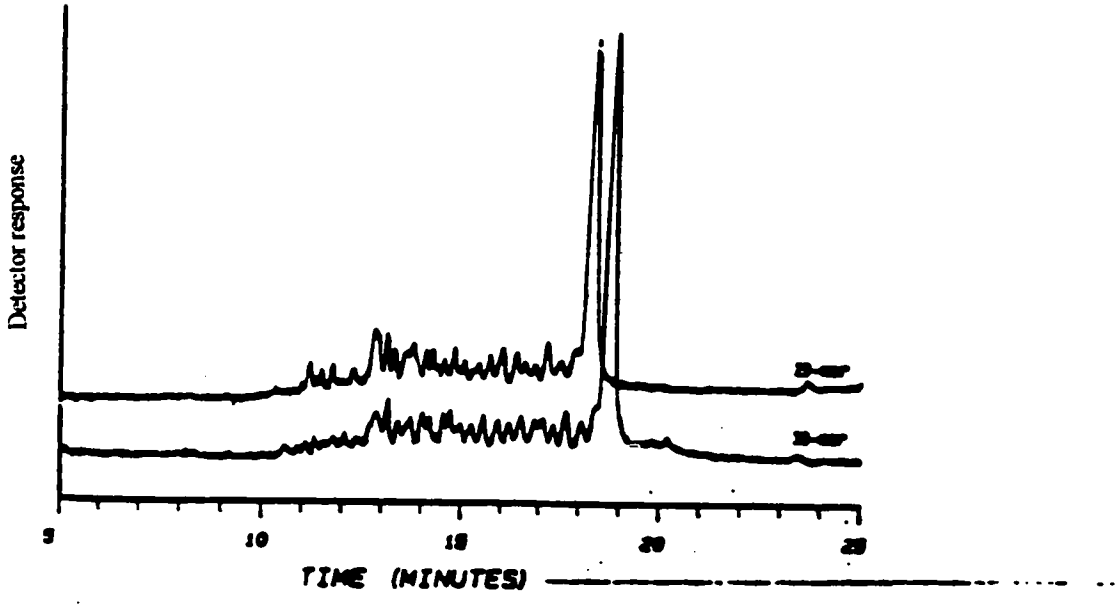


Figure 28. Separation of crude synthetic oligonucleotides, upper trace, 29-mer: 5'GCCCACCACCACAAGCTTGTATTCTGTCA3'; lower trace, 30-mer: 5'ATGACAGAATACAAGCTTGTGGTGGTGGGC3'[23]
Separation conditions: Field 250 V/cm; current 4.8 μA; injection 4000 Vs; other conditions same as Figure27..

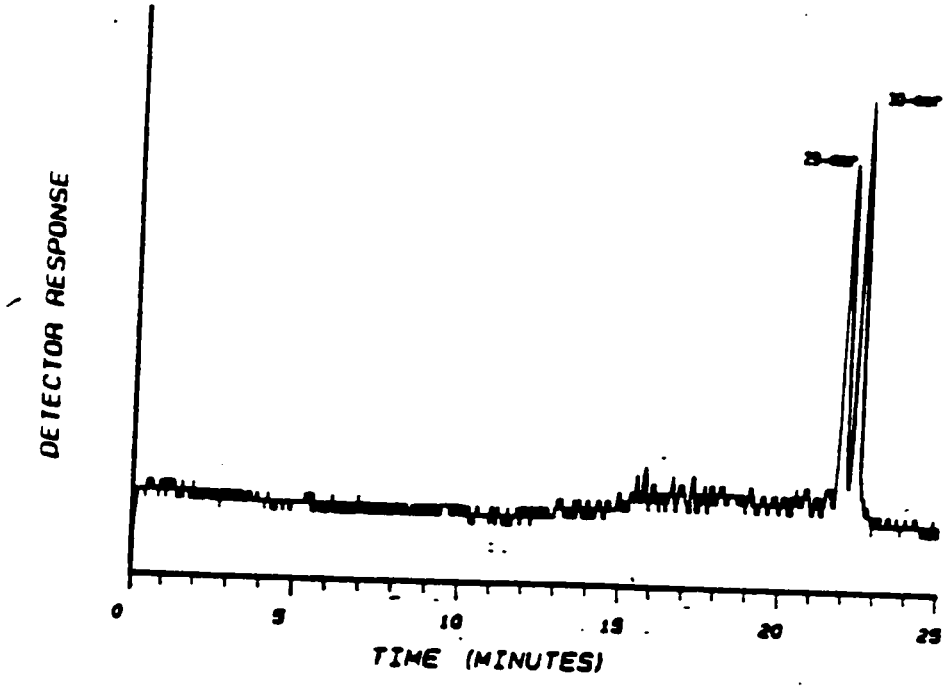


Figure 29. Separation of a mixture of two oligonucleotides[23]
Separation conditions: Same as Figure 27.

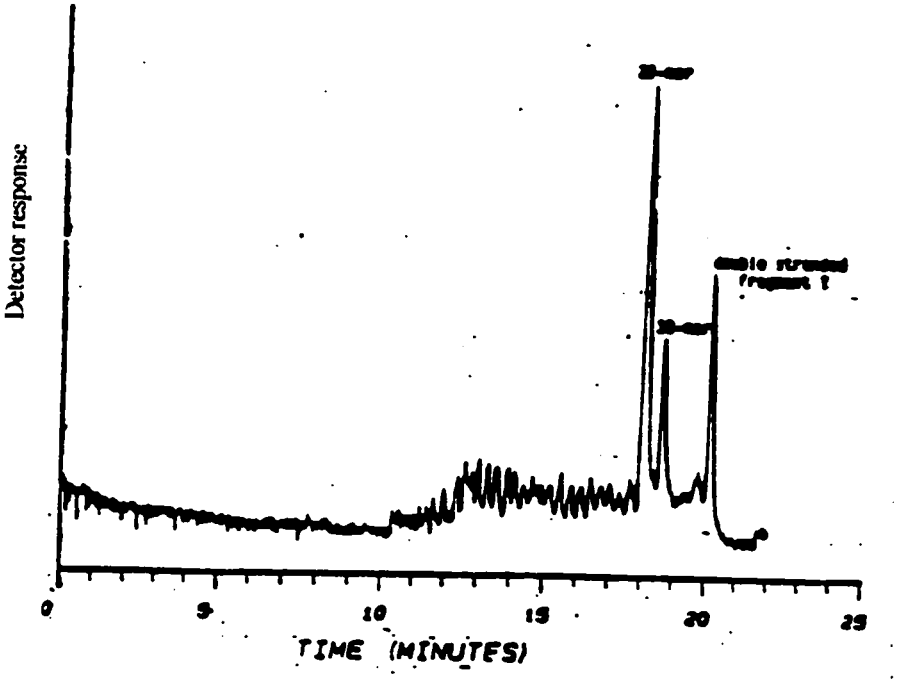


Figure 30. Separation of a mixture of two crude oligonucleotides, 29-mer: 5'GCCCACCACCACAAGCTTGTATTCTGTCA3', and 30-mer: 5'ATGACAGAATACAAGCTTGTGGTGGTGGGC3'[23]
Separation conditions: Same as Figure 27.

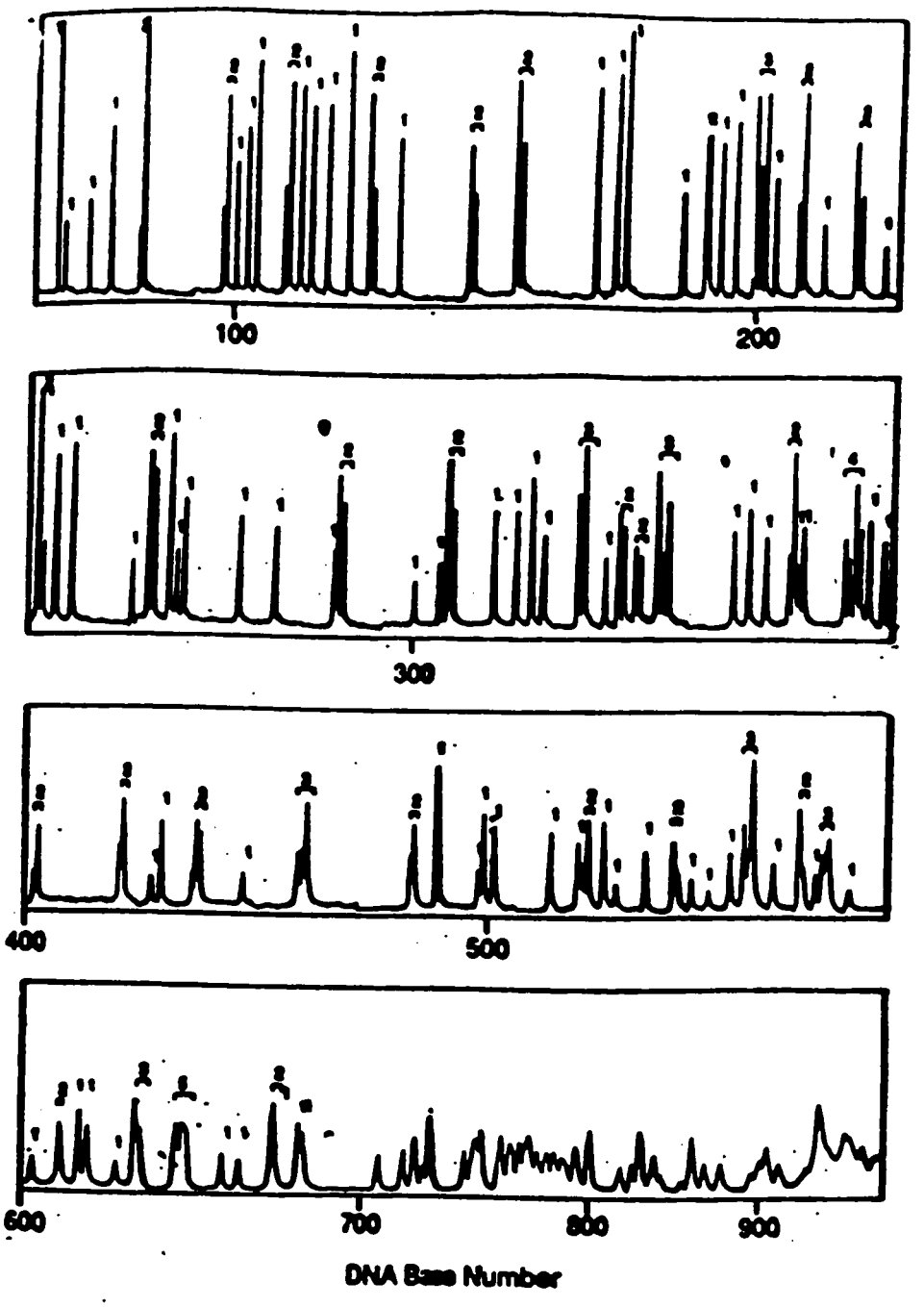


Figure 31. Pattern of DNA fragments (cytosine-reaction products) with base numbers from about 60-1000[30]

Injection of 0.4 pmole/μL of reaction products by applying an electric field of 150 V/cm for 15 s. Migration distance: 200 cm; capillary: 0.1 mm ID; gel concentration 3 % T, 5% C; field strength applied during electrophoresis: 170 V/cm.

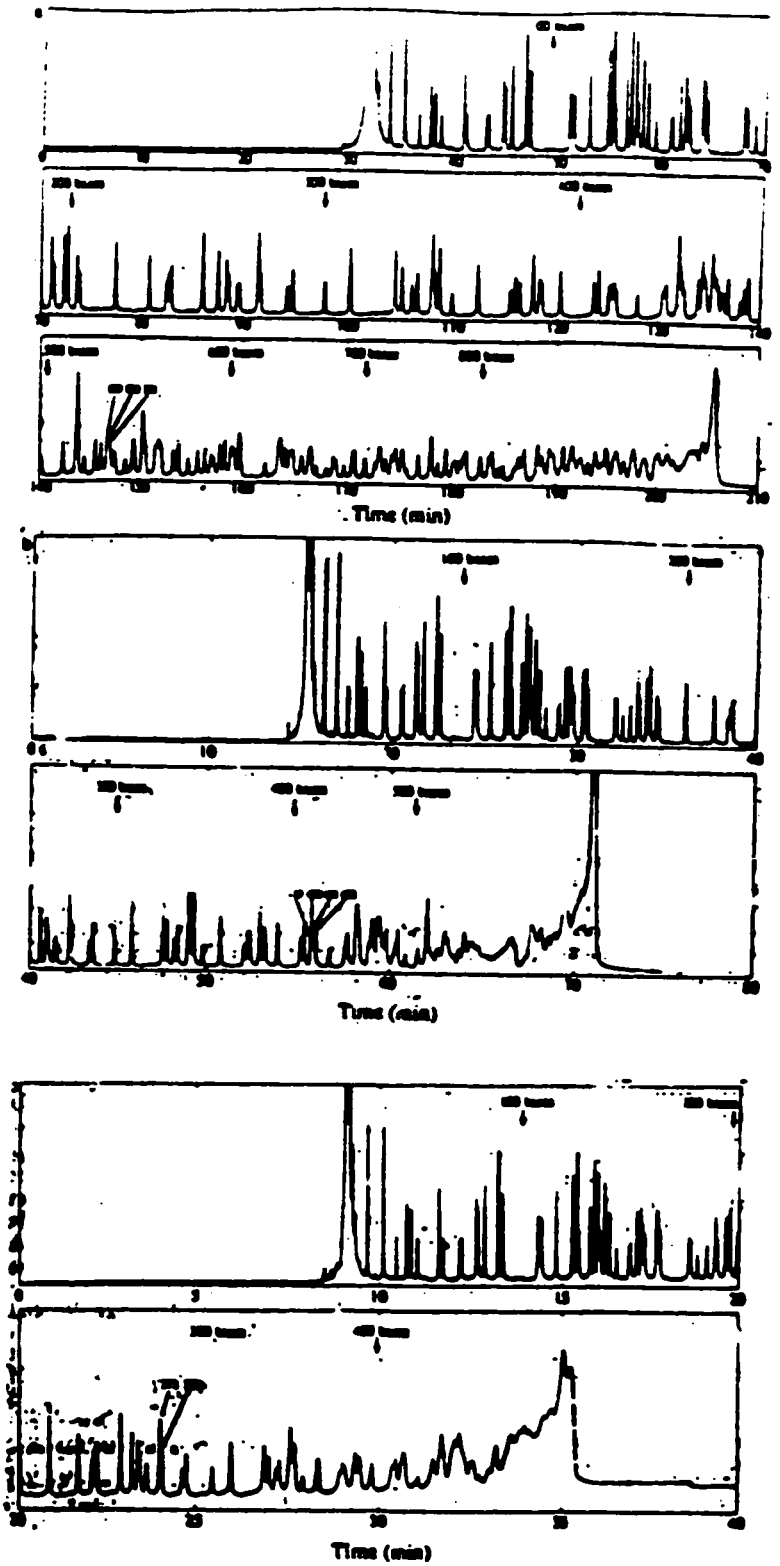


Figure 32. CGE patterns of DNA fragments in a 4% T, 5% C polyacrylamide gel. Electrophoresis at different electric field strengths: a) 100 V/cm, b) 200 V/cm, c) 300 V/cm[29]

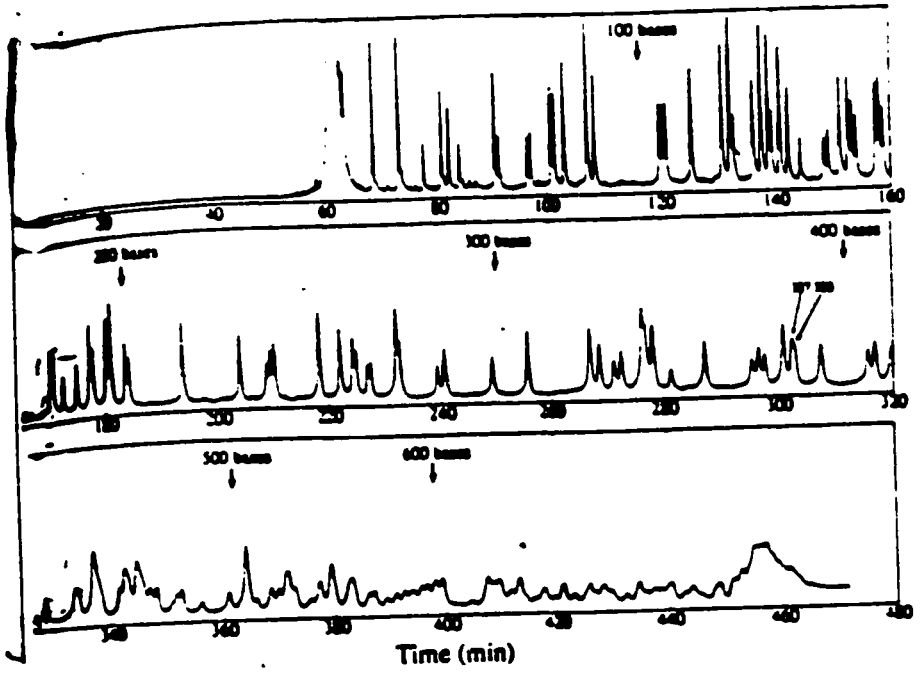


Figure 33. CGE patterns of DNA fragments with a 9 % T, 0% C polyacrylamide matrix at 100 V/cm[29]

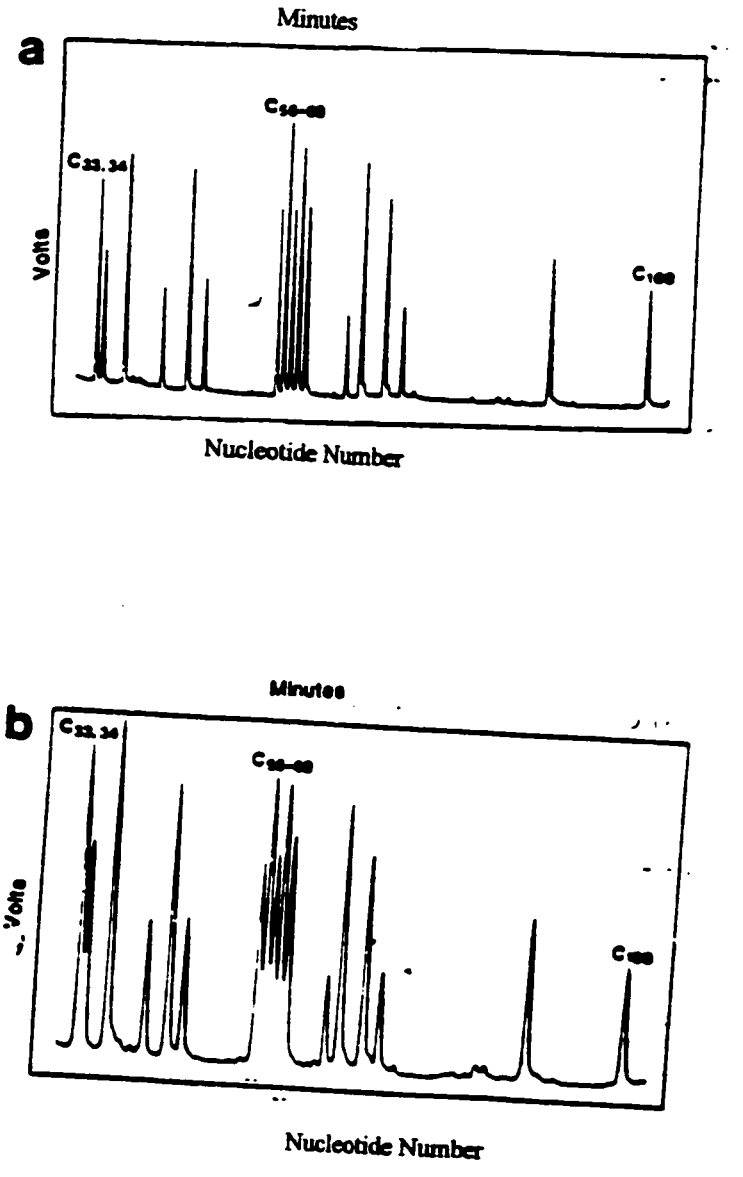


Figure 34. Comparison of capillary (a) and SGE (b) of a DNA sequencing reaction.[33] The peak labelled C100 took 54 min to reach the detector in a) and150 min in b). The abscissa represents time after sample introduction, and is displayed in minutes as well as corresponding oligonucleotide length; fluorescent signal is plotted on the ordinate as voltage.

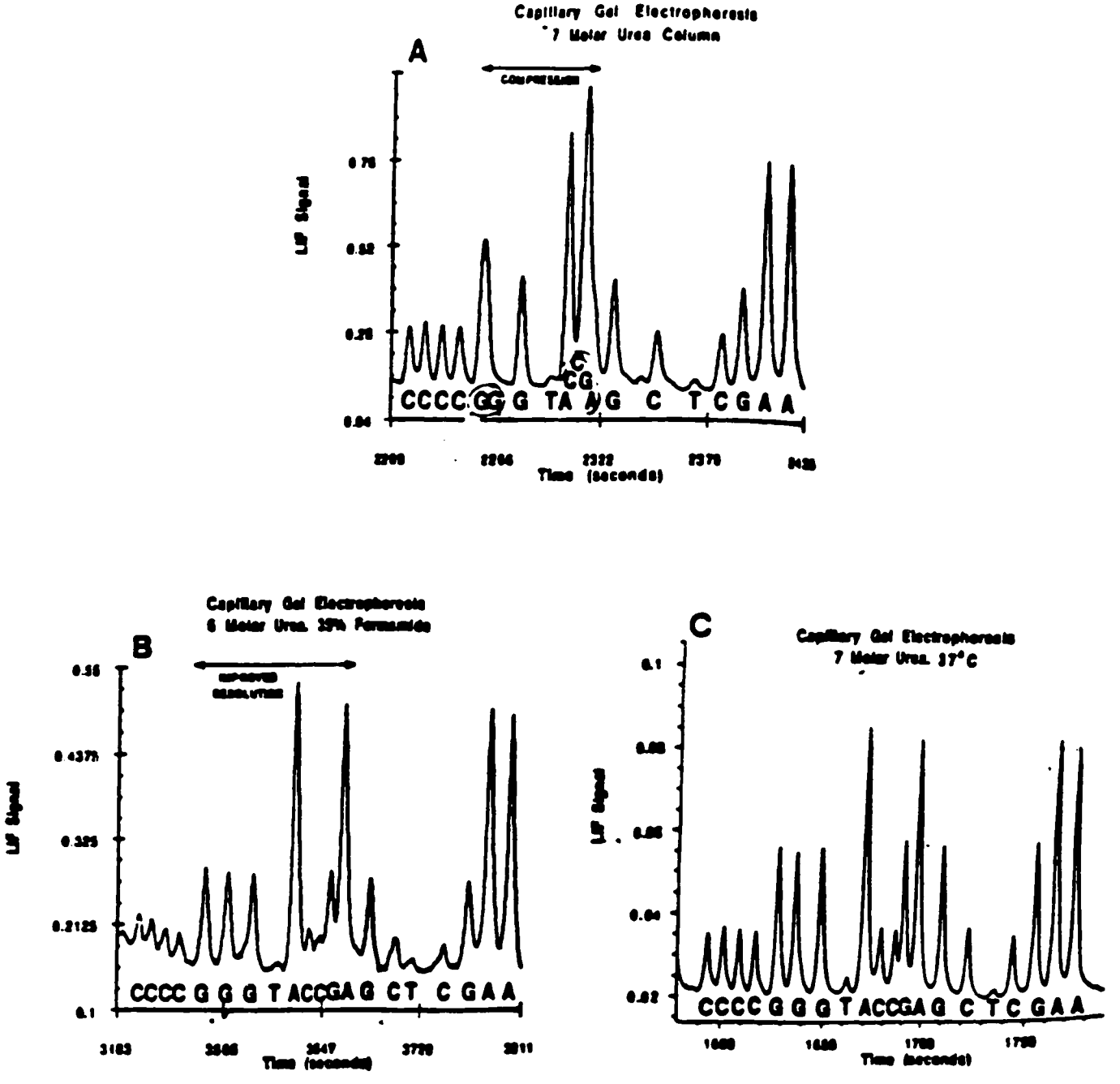


Figure 35. HPCE Electropherograms showing the separation of fluorescently labeled DNA fragments[33].

35a) shows the control, 35b) shows the reduction of compression using formamide and 35c) shows separation at 37 ^{0}C

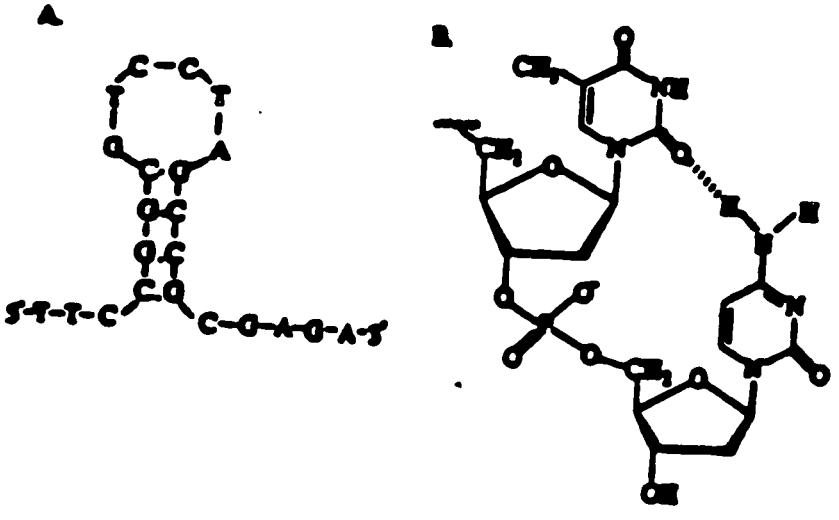


Figure 36. Possible secondary structures: (A) hairpin loop in a single-stranded oligonucleotide and (B) hydrogen bonding between two adjacent nucleotides[34]

	Surface area (mm^2)	Volume (μl)	Surface-to volume ratio
Slab gel (14 × 11.5 × 0.15 cm)	32200	24150	1.3
20 mm i.d. capillary	35.81	0.179	200
50 mm i.d. capillary	89.53	1.119	80
75 mm i.d. capillary	134.3	2.518	53
100 mm i.d. capillary	179.1	4.477	40
200 mm i.d. capillary	358.1	17.907	20

Table I. A comparison of surface-to-volume ratios for an analytical slab gel and a 57 cm capillary having a varied internal diameter[1]

Capillary coating	Purpose	Supplier and product name
Neutral Polymer	Reduce electroosmotic flow; prevent analyte-wall interactions	Beckman eCAP dsDNA 1000; eCAP Neutral/Protein; Bio-Rad BioCap linear polyacrylamide Hewlett-Packard poly-vinyl alcohol; Isco CE-200; J & W Scientific-DBWax, DB-17; Phenomenex CZESep-200/Glycerol; Supelco P150, P175; Scientific Resources PEG-100
Anionic	Stabilize EOF to improve reproducibility	MetaChem Technologies Microcap 2; Phenomenex CZESep-300/Sulfonic
Cationic	Reverse EOF	Beckman eCAP Amine; MetaChem Technologies Microcap 1. Supelco CElect Amine
Hydrophobic	Prevent analyte-wall interactions	Isco CE-100 (C18); Phenomenex CZESep-100/C18; Scientific Resources CE-200HS. Supelco H50, H75, H150, H175, H250, H275 J & W Scientific-DB-1, DB-5, DB-17
Detergent	Reduce analyte-wall interactions	Isco CE-300; Phenomenex CZESep-200/C-18
Chiral	Separation of enantiomers	Beckman eCAP Chiral Kit
Gel	Separation by size	Beckman eCAP SDS 14-200, eCAP dsDNA 20000 Kit; eCAP dsDNA 1000 Kit, eCAP ssDNA 100 Kit. Bio-Rad CE-SDS Protein Kit (entangled polymer for uncoated capillary); PCR product analysis Kit; Dionex NucleoPhor SB1 5 Kit. J & W Scientific mPAGE (3 or 5% polyacrylamide); Perkin-Elmer/Applied Biosystems ProSort protein analysis (physical gel); DNA Fragment analysis kit (replaceable matrix). Microgel nucleotide analysis kit (polyacrylamide gel); SRI MicroSolve CE 3%T 3%C or 5%T 5%C (polyacrylamide); Thermo Separation Products DNA Kit

Table II. Commercially available coated capillaries that can be used with any instrumentation[1]

Source	Wavelength (nm)
Lamp	
Deuterium	190–400 (continuous)
Xenon	190–2000 (continuous)
Hg-Xenon	190–2000 (continuous)
Hg-low pressure	200–550 (line)
Laser	
Ar-ion—air cooled	457. 472. 476. 488. 496. 501. 514
Full frame and υ-doubled	Above visible and 275. 300. 305. 333. 351. 364. 385. 229. 238. 244. 248. 257
Ar/Kr	Above visible and 350–360. 521. 531. 568. 647. 752.
HeCd	325. 354. and 442
Excimer	
XeCl	308
KrF	248
Nitrogen	337
Nitrogen pumped dye	360–950 (tunable)
Solid State	
YAG (υ-doubled)	532
YAG (υ-quadrupled)	266
Diode Lasers	620-infrared

Table III. Wavelength range of broad-band sources and lasers[1]

Polymer	Concentration	Application
Crosslinked polymers		
polyacrylamide/bis-acrylamide	2–6 % T, 3 – 6 % C	• Oligonucleotides, DNA sequencing, • Native and SDS-bound proteins
Linear polymers		
Polyacrylamide	< 0.1–6 %	• Restriction fragments
Hydroxylalkyl cellulose, polyvinyl alcohol, dextran	6 –15 %	• Oligonucleotides, DNA sequencing, proteins
Agarose	0.05 –1.2 %	• Restriction fragments • Proteins

Table IV. Polymer matrices for CGE[2]

Name	pK_a
Phosphate	2.12 (pK_{a_1})
Citrate	3.06 (pK_{a_1})
Formate	3.75
Succinate	4.19 (pK_{a_1})
Citrate	4.74 (pK_{a_2})
Acetate	4.75
Citrate	5.40 (pK_{a_3})
Succinate	5.57 (pK_{a_2})
MES	6.15
ADA	6.60
BIS-TRIS propane	6.80
PIPES	6.80
ACES	6.90
MOPSO	6.90
Imidazole	7.00
MOPS	7.20
Phosphate	7.21 (pK_{a_2})
TES	7.50
HEPES	7.55
HEPPS	8.00
TRICINE	8.15
Glycine amide, hydrochloride	8.20
Glycylglycine	8.25
TRIS	8.30
BICINE	8.35
Morpholine	8.49
Borate	9.24
CHES	9.50
CHAPSO	9.60
CAPS	10.40
Phosphate	12.32 (pK_{a_3})

Table V. Commonly used buffer[2]

Additive	Example	Function
Inorganic salts	NaCl, $CaCl_2$, K_2SO_4	Modification of EOF: protein conformational changes, protein hydration
Organic solvents	methanol, acetonitrile, ethylene glycol	Modification of EOF, analyte solubilization, analyte solvation
Organic additives	urea	Modification of EOF: protein solubilization: denaturation of oligonucleotides
Sulfonic acids	hexane, heptane, octane, or nonane analogs	Analyte ion-pairing: hydrophobic interaction
Divalent amines	diaminoalkanes: hexamethonium bromide: decamethonium bromide	Modification of EOF: charge neutralization: analyte interaction
Cationic surfactants	dodecyltrimethylammonium bromide (DTAB): cetyltrimethylammonium bromide (CTAB): tetradecyltrimethylammonium chloride (TTAC)	Charge reversal on capillary wall: hydrophobic interaction
Cellulose derivatives	hydroxyethyl cellulose: methyl cellulose: hydroxypropyl methylcellulose	Reduce EOF: provide sieving medium

Table VI. Common buffer additives in CE and their effects[1]

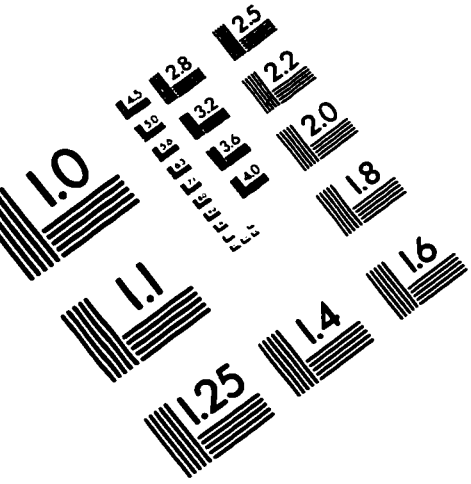
Mode	Basis of separation
Capillary zone electrophoresis	Free solution mobility
Micellar electrokinetic chromatography	Hydrophobic/ ionic interactions with micelle
Capillary gel electrophoresis	Size and charge
Isoelectric focusing	Isoelectric point
Isotachophoresis	Moving boundaries

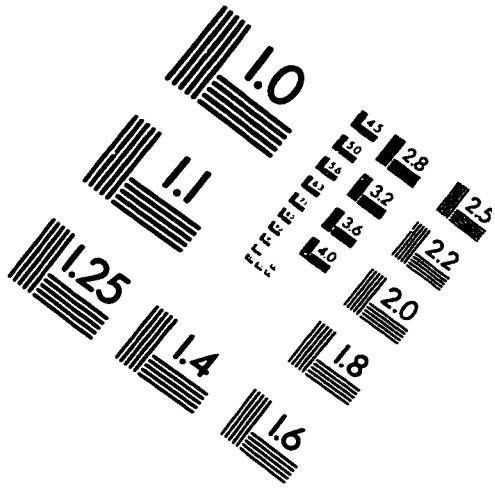
Table VII. Modes of CE[2]

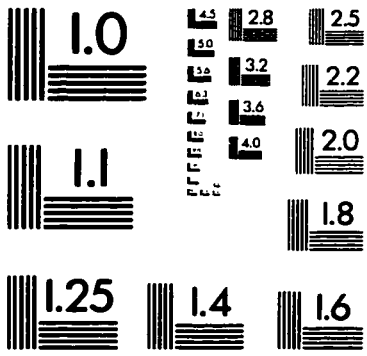
	Capillary gel		Slab gel	
Migration distance (cm)	200	300	35	30
Gel concentration (% T)	3	4	3	4.5
Electric field (V/cm)	170	70	190	35
Maximum separable base number (*Nm*)	680	800	420	450
Detection time (*Tm*, min)	585	2647	60	350

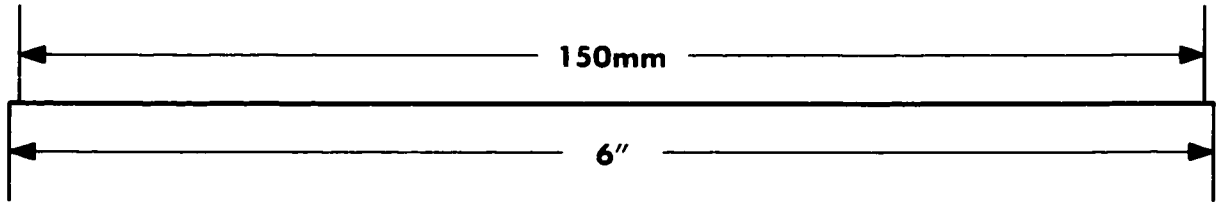
Table VIII. Comparison of Nm and Tm[29]

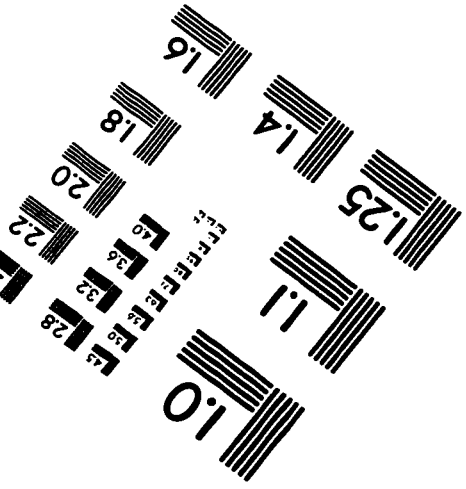
IMAGE EVALUATION TEST TARGET (QA-3)











APPLIED IMAGE, Inc
1653 East Main Street
Rochester, NY 14609 USA
Phone: 716/482-0300
Fax: 716/288-5989